

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 30, 2004, 11:16:19 ; Search time 9275 Seconds
(without alignments)
11164.044 Million cell updates/sec
Title: US-09-765-271-55
Perfect score: 2389
Sequence: 1 TTCTTACAGTGGACTGT.....TAAGTAAGAAAAATAAAC 2389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 90 summaries

Database :

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	6	AR120265 Sequence
2	2388	100.0	2389	6	AR340956 Sequence
3	2388	100.0	2389	6	BD063274 Streptoco
4	2388	100.0	2406	6	AX569139 Sequence
5	2388	100.0	2451	6	BD268048 Streptoco
6	2388	100.0	2451	6	AR344447 Sequence
7	2388	100.0	2541	1	AR291695 Streptoco
8	2388	100.0	8195	6	AR218862 Sequence
9	2388	100.0	8195	6	BD003774 Polynucle
10	2388	100.0	10256	1	AE007418 Streptoco
11	2388	100.0	349980	6	AX571763 Sequence
12	2365.6	99.0	10320	1	AE008479 Streptoco
13	2328.2	97.5	20035	2	SPNEU1915
14	1434.4	60.0	2535	1	AF340221 Streptoco
15	1374.2	57.5	2523	6	BD263585 Novel str
16	1374.2	57.5	2523	6	AX343072 Sequence
17	1374.2	57.5	2647	6	BD263590 Novel str
18	1374.2	57.5	2647	6	AX343073 Sequence
19	1011.8	42.4	2478	6	BD229972 Human com
20	1003.8	42.0	2481	6	BD264091 Streptoco
21	993	41.6	2457	6	AX569137 Sequence
22	991	41.5	2457	1	AF318954 Streptoco
23	991	41.5	2531	6	BD268049 Sequence
24	991	41.5	2531	6	AR344448 Sequence
25	990	41.4	2517	1	AF318955 Streptoco
26	990	41.4	2517	6	AX568778 Sequence
27	990	41.4	2531	6	BD268046 Streptoco
28	990	41.4	2531	6	AR344445 Sequence
29	990	41.4	11931	1	AE007403 Streptoco
30	990	41.4	349980	6	AX571762 Sequence
31	988.4	41.4	232807	2	SPNEU1901
32	987.6	41.3	2290	6	AR120270 Sequence
33	987.6	41.3	2290	6	AR340961 Sequence
34	987.6	41.3	2290	6	BD063279 Streptoco
35	980.8	41.1	2639	6	BD263591 Novel str
36	980.8	41.1	2639	6	AX343074 Sequence
37	962.8	40.3	12372	1	AE008464 Streptoco
38	869.4	36.4	2166	12	AF340222 Synthetic
39	835.8	35.0	75874	2	SPNEU1907
40	674	28.2	702	2	SPNEU1929
41	653.8	27.4	2359	6	AR219011 Sequence
42	653.8	27.4	2359	6	BD003923 Polynucle
43	548.4	23.0	3171	6	BD263631 Novel str
44	385.4	16.1	1342	6	AR120328 Sequence
45	385.4	16.1	1342	6	AR341019 Sequence
46	385.4	16.1	1342	6	BD063337 Streptoco
47	385.4	16.1	1455	6	BD263587 Novel str
48	385.4	16.1	1455	6	BD264131 Streptoco
49	385.4	16.1	1455	6	BD268047 Streptoco
50	385.4	16.1	1455	6	AR344446 Sequence
51	385.4	16.1	1455	6	BD223895 Nucleic a
52	385.4	16.1	3117	6	AX568780 Sequence
53	385.4	16.1	3120	1	AF318956 Streptoco
54	385.4	16.1	3120	6	BD263584 Novel str
55	385.4	16.1	3120	6	AX343070 Sequence
56	385.4	16.1	5048	6	BD263589 Novel str
57	385.4	16.1	5048	6	AX343071 Sequence
58	385.4	16.1	6867	6	AR218960 Sequence
59	385.4	16.1	6867	6	BD003872 Polynucle
60	381.4	16.0	973	6	AR219123 Sequence
61	381.4	16.0	973	6	BD004035 Polynucle
62	347	14.5	2528	6	BD263592 Novel str
63	347	14.5	2528	6	AX343078 Sequence
64	253.2	10.6	10312	1	AE007404 Streptoco
65	247.4	10.4	504	6	BD229970 Human com

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

C	66	247.4	10.4	504	6	BD229971	Human com
	67	243.4	10.2	1684	6	AR219026	Sequence
	68	243.4	10.2	1684	6	BD003938	Polynucle
C	69	219.6	9.2	75248	6	AX602204	Sequence
C	70	219.6	9.2	98050	1	SAG766854	Sequence
	71	214.6	9.0	2478	6	AX088372	Sequence
C	72	214.6	9.0	10825	1	AE006623	Streptoco
	73	213	8.9	2478	1	AE0073859	Streptoco
C	74	213	8.9	10618	1	AE010110	Streptoco
	75	211.4	8.8	1146	6	BD263932	Nucleic a
	76	211.4	8.8	1910	1	SAG290952	Streptoco
	77	211.4	8.8	2469	6	BD263929	Nucleic a
	78	211.4	8.8	2469	6	AX088376	Sequence
	79	211.4	8.8	2469	6	AX607520	Sequence
	80	211.4	8.8	3501	1	AF062533	Streptoco
C	81	211.4	8.8	11617	6	AX602163	Sequence
	82	211.4	8.8	21414	1	AE014248	Streptoco
C	83	211.4	8.8	199050	1	SAG766850	Streptoco
	84	209.8	8.8	2469	6	BD263632	Novel str
	85	209.8	8.8	5215	6	BD136831	Group B s
	86	206.6	8.6	2472	6	BD263633	Novel str
C	87	206	8.6	21397	1	AE014279	Streptoco
C	88	205	8.6	50354	1	AE014169	Streptoco
C	89	205	8.6	323825	1	AP005146	Streptoco
	90	201.6	8.4	1446	6	AX608407	Sequence

ALIGNMENTS

RESULT 1	AR120265	AR120265	2389 bp	DNA	Linear	PAT 16-MAY-2001
LOCUS	Sequence	55 from patent	US 6159469			
DEFINITION	AR120265	Sequence	55 from patent	US 6159469		
ACCESSION	AR120265	Sequence	55 from patent	US 6159469		
VERSION	AR120265.1	GI:14103841				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 2389)					
AUTHORS	Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.					
TITLE	Streptococcus pneumoniae antigens and vaccines					
JOURNAL	Patent: US 6159469-A 55 12-DEC-2000;					
FEATURES	Location/Qualifiers					
Source	1..2389					
	/organism="unknown"					
	/mol_type="unassigned DNA"					
ORIGIN						
Query Match	100.0%; Score 2388; DB 6; Length 2389;					
Best Local Similarity	100.0%; Pred. No. 0;					
Matches 2389; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA	60			
Db	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA	60			
QY	61	TATAGATGAAACACAGCAGCAGCAAAACCGAGATTTGACTCCTGATGAGGTTAGCAA	120			
Db	61	TATAGATGAAACACAGCAGCAGCAAAACCGAGATTTGACTCCTGATGAGGTTAGCAA	120			
QY	121	CGGTGAAGGAATCAATGCTGAGCAATCGTCAATAGATAACAGACCAAGGCTATGTCAC	180			
Db	121	CGGTGAAGGAATCAATGCTGAGCAATCGTCAATAGATAACAGACCAAGGCTATGTCAC	180			
QY	181	TTACATGCGGACCACTATCAATTAATGATGATGAGGTTCCCTATGACGCTATCATCAG	240			
Db	181	TTACATGCGGACCACTATCAATTAATGATGATGAGGTTCCCTATGACGCTATCATCAG	240			
QY	241	TGAAGAATTACTCATGAAAGATCCAACTATAGCTAAAGATGAGGATATGTTAATCA	300			

241	TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA	300
301	GGTCAAGGTTGGATATGTTATCAAGCTAGATGAAATACTATCTTTTACCTTAAGATGC	360
301	GGTCAAGGTTGGATATGTTATCAAGCTAGATGAAATACTATCTTTTACCTTAAGATGC	360
361	TGCCCAACGGGATAACGCTCGTACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	420
361	TGCCCAACGGGATAACGCTCGTACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	420
421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTTGGCCTTGGCAGCTTCGCA	480
421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTTGGCCTTGGCAGCTTCGCA	480
481	AGGACGCTATCTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC	540
481	AGGACGCTATCTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC	540
541	TGCTGATGCTTATATCTCTCATGAGATCATTAACCATTTACATTCCTAAGAAATGAGTT	600
541	TGCTGATGCTTATATCTCTCATGAGATCATTAACCATTTACATTCCTAAGAAATGAGTT	600
601	ATCAGTAGGAGTTGGCTGCTGAGAACCTTCTATCTGTCGAGGAAATCTGCAAA	660
601	ATCAGTAGGAGTTGGCTGCTGAGAACCTTCTATCTGTCGAGGAAATCTGCAAA	660
661	TTCAAGAACCTATCGCCGACAAATAGCGATACACTTCAAGAACAACTGGGTACTCTTC	720
661	TTCAAGAACCTATCGCCGACAAATAGCGATACACTTCAAGAACAACTGGGTACTCTTC	720
721	TGTAAGCAATCCAGGAATCTACAAATACTAACACAAAGCAACAACTAACAGTCA	780
721	TGTAAGCAATCCAGGAATCTACAAATACTAACACAAAGCAACAACTAACAGTCA	780
781	AGCAAGTCAAGATATGACATTTAGTCTCTTTGAACAGCTCTACAACTGCTTTGAG	840
781	AGCAAGTCAAGATATGACATTTAGTCTCTTTGAACAGCTCTACAACTGCTTTGAG	840
841	TCAACGACATGATGATCTGATGGCTCTCTTTGATCCAGCACAACTCAGTCTGCAAC	900
841	TCAACGACATGATGATCTGATGGCTCTCTTTGATCCAGCACAACTCAGTCTGCAAC	900
901	AGCTAGAGTGTGTCAGTCCACACGAGATCATTAACCATTTCTCTCTTACTCTCAAA	960
901	AGCTAGAGTGTGTCAGTCCACACGAGATCATTAACCATTTCTCTCTTACTCTCAAA	960
961	GTCGAAATTTGGAAGAACGAACTGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG	1020
961	GTCGAAATTTGGAAGAACGAACTGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG	1020
1021	GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGS	1080
1021	GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGS	1080
1081	CCGCAACCTGCAACCAAAATCTTTAAATAGACTCAAATCTCTTTGGTTAGTCAGTGT	1140
1081	CCGCAACCTGCAACCAAAATCTTTAAATAGACTCAAATCTCTTTGGTTAGTCAGTGT	1140
1141	ACGAAAGTTGGGAGGATATGATTCGAAAGAAAGGCACTCTCTGTTATGCTTTTC	1200
1141	ACGAAAGTTGGGAGGATATGATTCGAAAGAAAGGCACTCTCTGTTATGCTTTTC	1200
1201	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTATCAAAAACAGAGAG	1260
1201	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTATCAAAAACAGAGAG	1260
1261	TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA	1320
1261	TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA	1320
1321	TGATAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTTGTTGNAATAAGGCTCG	1380
1321	TGATAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTTGTTGNAATAAGGCTCG	1380

QY	1381	TAATTCGATTTCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGATGACTAA	1440
Db	1381	TAATTCGATTTCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGATGACTAA	1440
QY	1441	TAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCAAAATTAACCCATCAGAGCG	1500
Db	1441	TAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCAAAATTAACCCATCAGAGCG	1500
QY	1501	ACTTGGCAAAACAAATTCCTAAATTTAGTATATCTGAAGACGAAGTTCGTATTGCTCAAT	1560
Db	1501	ACTTGGCAAAACAAATTCCTAAATTTAGTATATCTGAAGACGAAGTTCGTATTGCTCAAT	1560
QY	1561	AGCTGATAGTATACAAACGTCAGATGTTTATTTTGTGAACATGATATTAATCAGTGA	1620
Db	1561	AGCTGATAGTATACAAACGTCAGATGTTTATTTTGTGAACATGATATTAATCAGTGA	1620
QY	1621	TGAAGGATGATCATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAGAATAG	1680
Db	1621	TGAAGGATGATCATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAGAATAG	1680
QY	1681	CCTTTCTGATAAGGAAAAGTTGCAGCTCAAGCCTATATAAGAAAAGGTTATCTTACC	1740
Db	1681	CCTTTCTGATAAGGAAAAGTTGCAGCTCAAGCCTATATAAGAAAAGGTTATCTTACC	1740
QY	1741	TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA	1800
Db	1741	TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA	1800
QY	1801	CAATCGTGTGAAGGGGAAAAACGAATTCCTACTCGTTCGATTCATATATGTTGAGCA	1860
Db	1801	CAATCGTGTGAAGGGGAAAAACGAATTCCTACTCGTTCGATTCATATATGTTGAGCA	1860
QY	1861	TACAGTTGAGTTAAAAACGGTAATTTGATTAATTCCTCATAGGATCATTAACCAATAT	1920
Db	1861	TACAGTTGAGTTAAAAACGGTAATTTGATTAATTCCTCATAGGATCATTAACCAATAT	1920
QY	1921	TAAATTTGCTTGTGTTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAGA	1980
Db	1921	TAAATTTGCTTGTGTTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAGA	1980
QY	1981	TTTGTGTTGCGACGATTAAGTACTAGTAGAACACCCCTGACGAACCTCCACATCTTAATGA	2040
Db	1981	TTTGTGTTGCGACGATTAAGTACTAGTAGAACACCCCTGACGAACCTCCACATCTTAATGA	2040
QY	2041	TGGATGGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	2100
Db	2041	TGGATGGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	2100
QY	2101	TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC	2160
Db	2101	TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC	2160
QY	2161	TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTGGCAA	2220
Db	2161	TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTGGCAA	2220
QY	2221	AGTAACCGATTCAGTCTGAAGGCAATGCAACGAAACCTTAGCTGTTTACGAAATAA	2280
Db	2221	AGTAACCGATTCAGTCTGAAGGCAATGCAACGAAACCTTAGCTGTTTACGAAATAA	2280
QY	2281	TTTGACTCTTCAAAATTTAGGATACAAATAGTATCATGGCAAGCAGAAATTAATCTTGC	2340
Db	2281	TTTGACTCTTCAAAATTTAGGATACAAATAGTATCATGGCAAGCAGAAATTAATCTTGC	2340
QY	2341	GTGTTAAAGGAAGTAATCTTCACTCTGTGAAGTAAAGGAAAAATAAAC	2389
Db	2341	GTGTTAAAGGAAGTAATCTTCACTCTGTGAAGTAAAGGAAAAATAAAC	2389

RESULT 2
AR340956
LOCUS

DEFINITION	Sequence 55 from patent US 6573082.
ACCESSION	AR340956
VERSION	AR340956.1 GI:33732935
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 2389) Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A. Streptococcus pneumoniae antigens and vaccines Patent: US 6573082-A 55 03-JUN-2003; Location/Qualifiers 1..2389 /organism="unknown" /mol_type="genomic DNA"
TITLE	
JOURNAL	
FEATURES	
ORIGIN	
Query Match	100.0%; Score 2388; DB 6; Length 2389;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2389; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 60
Db	1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 60
QY	61 TATAGATGGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
Db	61 TATAGATGGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
QY	121 GCGTGAAGGATCAATCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCCAC 180
Db	121 GCGTGAAGGATCAATCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCCAC 180
QY	181 TTCATATGGCGACCACTATCATTTACATTTGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db	181 TTCATATGGCGACCACTATCATTTACATTTGTAAGGTTCCCTTATGACGCTATCATCAG 240
QY	241 TGAAGAATTACTCATGAAGATCCAACTATAAGCTAAAAGATGAGGATTTGTTAATGA 300
Db	241 TGAAGAATTACTCATGAAGATCCAACTATAAGCTAAAAGATGAGGATTTGTTAATGA 300
QY	301 GGTCAAGGGTGATATGTTATCAAGGTAGTAGGAAATACCTATGTTTACCTTAAGGATGC 360
Db	301 GGTCAAGGGTGATATGTTATCAAGGTAGTAGGAAATACCTATGTTTACCTTAAGGATGC 360
QY	361 TGCCACGCGGATACGTCCTGACAAAAGAGGAAATCAATCGACAAAACAAAGAGCATAG 420
Db	361 TGCCACGCGGATACGTCCTGACAAAAGAGGAAATCAATCGACAAAACAAAGAGCATAG 420
QY	421 TCAACATCTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGCTTGGCAAGTTCGCA 480
Db	421 TCAACATCTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGCTTGGCAAGTTCGCA 480
QY	481 AGGACGCTTACTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db	481 AGGACGCTTACTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY	541 TGGTATGCTTTATATCGTTCCCTCATGGAGATCATTAACATTAATTCCTAAGAATGAGTT 600
Db	541 TGGTATGCTTTATATCGTTCCCTCATGGAGATCATTAACATTAATTCCTAAGAATGAGTT 600
QY	601 ATGAGCTAGCGAGTTGGTGTCTGAGAAAGCCTTCCCTATCTGGTTCGAGGAAATCTGTCAA 660
Db	601 ATGAGCTAGCGAGTTGGTGTCTGAGAAAGCCTTCCCTATCTGGTTCGAGGAAATCTGTCAA 660
QY	661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAAACAACTGGTACCTTC 720
Db	661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAAACAACTGGTACCTTC 720
QY	721 TGTAAAGCAATCCAGGAATCAAAATACTTAACAAGCAACAAACAGCAACACTTAACAGTCA 780
Db	721 TGTAAAGCAATCCAGGAATCAAAATACTTAACAAGCAACAAACAGCAACACTTAACAGTCA 780

781 AGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 840
 Db |||||
 781 AGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 840
 Qy |||||
 841 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCACAAAGTGAAC 900
 Db |||||
 841 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCACAAAGTGAAC 900
 Qy |||||
 901 AGCTAGAGTGTGCAAGTGCACACGAGATCATTTACCACTTCCCTTACTCTCAAAAT 960
 Db |||||
 901 AGCTAGAGTGTGCAAGTGCACACGAGATCATTTACCACTTCCCTTACTCTCAAAAT 960
 Qy |||||
 961 GTCTGAATTTGGAAGAACCAATCGCTCGTATTTATCCCTTCCCTTCCCTTCAAAACATTTG 1020
 Db |||||
 961 GTCTGAATTTGGAAGAACCAATCGCTCGTATTTATCCCTTCCCTTCCCTTCAAAACATTTG 1020
 Qy |||||
 1021 GGTACAGATTTCAAGGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 Db |||||
 1021 GGTACAGATTTCAAGGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 Qy |||||
 1081 CCGGCACTGACCAAAATCTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
 Db |||||
 1081 CCGGCACTGACCAAAATCTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
 Qy |||||
 1141 ACGAAAGTTGGGGAAGGATATGATTTGAAAGAAAGGGGATCTCTCGTTATGCTTTTC 1200
 Db |||||
 1141 ACGAAAGTTGGGGAAGGATATGATTTGAAAGAAAGGGGATCTCTCGTTATGCTTTTC 1200
 Qy |||||
 1201 GAAAGATTTACCAATCTGAACTGTTAAATCTTGAAGAACGAGTTATCAAAACAGAGAG 1260
 Db |||||
 1201 GAAAGATTTACCAATCTGAACTGTTAAATCTTGAAGAACGAGTTATCAAAACAGAGAG 1260
 Qy |||||
 1261 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCGCTGACCAAAATTTTA 1320
 Db |||||
 1261 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCGCTGACCAAAATTTTA 1320
 Qy |||||
 1321 TGAATAAGCATATACTGTTAACTGAGGCTATAAAGGCTTGTGTTGNAATAAAGGCTG 1380
 Db |||||
 1321 TGAATAAGCATATACTGTTAACTGAGGCTATAAAGGCTTGTGTTGNAATAAAGGCTG 1380
 Qy |||||
 1381 TAATTCGTATTTCCAGGCTTAGACAAATTTAGAAAGCGTTGAATGATGAATCGACTAA 1440
 Db |||||
 1381 TAATTCGTATTTCCAGGCTTAGACAAATTTAGAAAGCGTTGAATGATGAATCGACTAA 1440
 Qy |||||
 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTTCTAGCACCAATACCCATCCAGAG 1500
 Db |||||
 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTTCTAGCACCAATACCCATCCAGAG 1500
 Qy |||||
 1501 ACTTGGCAACCAAAATTTCTCAATTTGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT 1560
 Db |||||
 1501 ACTTGGCAACCAAAATTTCTCAATTTGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT 1560
 Qy |||||
 1561 AGCTGATAGTATACAACTGATGTTTACATTTTGTATGAACATGATATATAGTGA 1620
 Db |||||
 1561 AGCTGATAGTATACAACTGATGTTTACATTTTGTATGAACATGATATATAGTGA 1620
 Qy |||||
 1621 TGAAGGAGATGATATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
 Db |||||
 1621 TGAAGGAGATGATATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
 Qy |||||
 1681 CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTTAAAGAAAAAGGTTATCTTACC 1740
 Db |||||
 1681 CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTTAAAGAAAAAGGTTATCTTACC 1740
 Qy |||||
 1741 TCCATCTCCAGAGGATGTTTAAAGCAATCAACTGAGATAGTGCAGCTATTATTA 1800
 Db |||||
 1741 TCCATCTCCAGAGGATGTTTAAAGCAATCAACTGAGATAGTGCAGCTATTATTA 1800
 Qy |||||
 1801 CAACTCGTGAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
 Db |||||
 1801 CAACTCGTGAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860

1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCAATACCATATAT 1920
 Db |||||
 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCAATACCATATAT 1920
 Qy |||||
 1921 TAAATTTGCTTGGTTTGGTATGATCACAATCAAAAGCTCAAAATGGCTATACCTTGAAGA 1980
 Db |||||
 1921 TAAATTTGCTTGGTTTGGTATGATCACAATCAAAAGCTCAAAATGGCTATACCTTGAAGA 1980
 Qy |||||
 1981 TTTGTTTGGGAGATTAAGTACTAGTAGAACACCTGAGACGCTCCACATTTCTAATGA 2040
 Db |||||
 1981 TTTGTTTGGGAGATTAAGTACTAGTAGAACACCTGAGACGCTCCACATTTCTAATGA 2040
 Qy |||||
 2041 TGGATGGGCAATGCGCAGTGGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 Db |||||
 2041 TGGATGGGCAATGCGCAGTGGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 Qy |||||
 2101 TGAAGATTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC 2160
 Db |||||
 2101 TGAAGATTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC 2160
 Qy |||||
 2161 TCAAGTAGAGACTTGAAGAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTCGGAA 2220
 Db |||||
 2161 TCAAGTAGAGACTTGAAGAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTCGGAA 2220
 Qy |||||
 2221 AGTAACGGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
 Db |||||
 2221 AGTAACGGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
 Qy |||||
 2281 TTTGACTTTCAAAATTTATGATTAACAATAGTATCATGCGAGAGACGAAATAATCTTGC 2340
 Db |||||
 2281 TTTGACTTTCAAAATTTATGATTAACAATAGTATCATGCGAGAGACGAAATAATCTTGC 2340
 Qy |||||
 2341 GTTGTAAAAAGGAAGTAATCTTCTATCTGTAAGTAAAGAAAAATAAAC 2389
 Db |||||
 2341 GTTGTAAAAAGGAAGTAATCTTCTATCTGTAAGTAAAGAAAAATAAAC 2389

RESULT 3
 BD063274
 LOCUS 2389 bp DNA linear PAT 27-AUG-2002
 DEFINITION Streptococcus pneumoniae antigens and vaccines.
 ACCESSION BD063274
 VERSION BD063274.1 GI:22608877
 KEYWORDS JP 2001505415-A/28.
 SOURCE unidentified
 ORGANISM Streptococcus pneumoniae
 REFERENCE 1 (bases 1 to 2389)
 AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromockyj,A.
 TITLE Streptococcus pneumoniae antigens and vaccines
 JOURNAL Patent: JP 2001505415-A 28 24-APR-2001;
 HUMAN GENOME SCIENCES INC
 COMMENT PN JP 2001505415-A/28
 PD 24-APR-2001
 PF 30-OCT-1997 JP 1998520667
 PR 31-OCT-1996 US 60/029960
 PI CHARLES A KUNSCH,GIL H CHOI,SYDNOR L JOHNSON,ALEX HROMOCKYJ PC
 C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
 G01N33/569,
 CC GOIN33/68
 CC Strandedness: Double;
 CC Topology: Linear;
 PH Key Location/Qualifiers.
 FEATURES
 1. 2389
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 ORIGIN
 Query Match 100.0%; Score 2388; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
Db 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
QY 61 TATAGATGGAAACAAAGCGACGCAAAAAACGAGAAATTGACTCTCGATGAGGTTAGCAA 120
Db 61 TATAGATGGAAACAAAGCGACGCAAAAAACGAGAAATTGACTCTCGATGAGGTTAGCAA 120
QY 121 GCGTGAAGGAATCAATGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCCAC 180
Db 121 GCGTGAAGGAATCAATGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCCAC 180
QY 181 TTCACATGCGACCACTATCATTTACAAATGGAAGGTTCCCTATGACGCTATCATCAG 240
Db 181 TTCACATGCGACCACTATCATTTACAAATGGAAGGTTCCCTATGACGCTATCATCAG 240
QY 241 TGAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTGTTAAATGA 300
Db 241 TGAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTGTTAAATGA 300
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTTAAGGATGC 360
Db 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTTAAGGATGC 360
QY 361 TGCCACGCGGATAACGTCGTTACAAAAGAGAAATCAATCGACAAAAACAGAGCATAG 420
Db 361 TGCCACGCGGATAACGTCGTTACAAAAGAGAAATCAATCGACAAAAACAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGTGCTTTGGCACGTTGCA 480
Db 421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGTGCTTTGGCACGTTGCA 480
QY 481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTCTGTATCATAGAGATAC 540
Db 481 AGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTCTGTATCATAGAGATAC 540
QY 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCACTACATTCTCTAAGATGAGTT 600
Db 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTACCACTACATTCTCTAAGATGAGTT 600
QY 601 ATCAGCTAGCGAGTTGGCTGTGCAAGAGCCTTCCTATCTGGTCGAGGAAATCTGTCAA 660
Db 601 ATCAGCTAGCGAGTTGGCTGTGCAAGAGCCTTCCTATCTGGTCGAGGAAATCTGTCAA 660
QY 661 TTCAGAACCTTATCGCCAGCAAAATAGCGATAACACTTCAAGAACAACTGGTACCTTC 720
Db 661 TTCAGAACCTTATCGCCAGCAAAATAGCGATAACACTTCAAGAACAACTGGGTAACCTTC 720
QY 721 TGTAAAGCAATCCAGGAATCAAAATACTAACACAGCAACACAGCAACACTTAACAGTCA 780
Db 721 TGTAAAGCAATCCAGGAATCAAAATACTAACACAGCAACACAGCAACACTTAACAGTCA 780
QY 781 AGCAAGTCAAAAGTAATGACATTGATGCTCTTTGAAACAGCTCTACAAACTGCCTTTGAG 840
Db 781 AGCAAGTCAAAAGTAATGACATTGATGCTCTTTGAAACAGCTCTACAAACTGCCTTTGAG 840
QY 841 TCAAGCATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 900
Db 841 TCAAGCATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 900
QY 901 AGCTAGAGGTGTGAGTGGCCACAGGAGATCATTACCACTTCATCCCTTACTCTCAAT 960
Db 901 AGCTAGAGGTGTGAGTGGCCACAGGAGATCATTACCACTTCATCCCTTACTCTCAAT 960
QY 961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACATTG 1020
Db 961 GTCTGAATTGGAAGAACGAATCGCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACATTG 1020
QY 1021 GGTACCAATTCAGGCGCAGAAACCAAGTCACAAACCGACTCCGGAACCTAGTCCAGG 1080
Db 1021 GGTACCAATTCAGGCGCAGAAACCAAGTCACAAACCGACTCCGGAACCTAGTCCAGG 1080

QY 1081 CCCGCAACCTGCACAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1140
Db 1081 CCCGCAACCTGCACAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1140
QY 1141 ACGAAAGCTTGGGAAGGATATGATTGAAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1200
Db 1141 ACGAAAGCTTGGGAAGGATATGATTGAAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1200
QY 1201 GAAAGATTTACATCTGAAAACCTGTTTAAAAATCTTGAAGAGCAAGTTATCAAAACAAGAG 1260
Db 1201 GAAAGATTTACATCTGAAAACCTGTTTAAAAATCTTGAAGAGCAAGTTATCAAAACAAGAG 1260
QY 1261 TGTTTACACACTTTTAACTGCTTAAAAAGAAAATGCTGCTCTCTGAGCCAGAAATTTTA 1320
Db 1261 TGTTTACACACTTTTAACTGCTTAAAAAGAAAATGCTGCTCTCTGAGCCAGAAATTTTA 1320
QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380
Db 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380
QY 1381 TAAATCTGATTTCCAAAGCCTTAGACAAATTAATAGAACGCTTGAATGATGAATCACTAA 1440
Db 1381 TAAATCTGATTTCCAAAGCCTTAGACAAATTAATAGAACGCTTGAATGATGAATCACTAA 1440
QY 1441 TAAAGAAAAATTTGGTAGATGATTTATTGCGCATTCCTAGCACCAATTAACCCATCCAGAGC 1500
Db 1441 TAAAGAAAAATTTGGTAGATGATTTATTGCGCATTCCTAGCACCAATTAACCCATCCAGAGC 1500
QY 1501 ACTTGGCAAAACCAAATTTCTCAAAATTGAGTATACTGAAGACGAAAGTTGCTATTGCTCAAT 1560
Db 1501 ACTTGGCAAAACCAAATTTCTCAAAATTGAGTATACTGAAGACGAAAGTTGCTATTGCTCAAT 1560
QY 1561 AGCTGATAAGTATACAACTGATGATGTTTATGATGAAATGATGATGATGATGATGATGAT 1620
Db 1561 AGCTGATAAGTATACAACTGATGATGTTTATGATGAAATGATGATGATGATGATGATGAT 1620
QY 1621 TGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 TGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 CCTTCTGATAAGGAAAAAGTTGAGGCTCAAGCCTTATACTAAAGAAAAAGGTTATCCTACC 1740
Db 1681 CCTTCTGATAAGGAAAAAGTTGAGGCTCAAGCCTTATACTAAAGAAAAAGGTTATCCTACC 1740
QY 1741 TCCATCTCAGACGCGATGTTTAAAGGAAATCCTGAGATGATGATGATGATGATGATGAT 1800
Db 1741 TCCATCTCAGACGCGATGTTTAAAGGAAATCCTGAGATGATGATGATGATGATGATGAT 1800
QY 1801 CAATCGTGTGAAAGGGGAAAAACGAAATTCCTGCTGCTTCCATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAAGGGGAAAAACGAAATTCCTGCTGCTTCCATATATGTTGAGCA 1860
QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCAATTAACATAAT 1920
Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCAATTAACATAAT 1920
QY 1921 TAAATTTGCTTTGGTTGATGATCACAATCAAGAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1921 TAAATTTGCTTTGGTTGATGATCACAATCAAGAGCTCCAAATGGCTATACCTTGGAGA 1980
QY 1981 TTTGTTGCGAGCAATTAAGTACTAGTAGAACACCTCGACCAACCTCCACATTTCTTAATGA 2040
Db 1981 TTTGTTGCGAGCAATTAAGTACTAGTAGAACACCTCGACCAACCTCCACATTTCTTAATGA 2040
QY 2041 TGGATGGGGCAATCCAGTGAGCATGTTTGAAGCAAGAAACCAAGTGAAGATCCAAA 2100
Db 2041 TGGATGGGGCAATCCAGTGAGCATGTTTGAAGCAAGAAACCAAGTGAAGATCCAAA 2100
QY 2101 TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
Db 2101 TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGACGAGAAAGTTTGTGTCGAA 2220

```
Db 2161 TCAAGTAGAGCTGAAAGTAGAAGCCCACTCAAGAGCAGAGAGTTTGGCTTCGGA 2220
QY 2221 AGTAACGAGATTCTAGTCTGAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2280
Db 2221 AGTAACGAGATTCTAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2280
QY 2281 TTGTACTCTCAAAATTATGATACCAATAGTATCATGTCGAGAGCAGAAATTTACTTGC 2340
Db 2281 TTGTACTCTCAAAATTATGATACCAATAGTATCATGTCGAGAGCAGAAATTTACTTGC 2340
QY 2341 GTTGTAAAAGGAAGTAATCTTCACTCTGTAACTAAGGAAAAATAAAC 2389
Db 2341 GTTGTAAAAGGAAGTAATCTTCACTCTGTAACTAAGGAAAAATAAAC 2389

RESULT 4
AX569139
LOCUS 2406 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 2347 from Patent WO02077021.
ACCESSION AX569139
VERSION AX569139.1 GI:26002636
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Massignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2347 03-OCT-2002;
Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1. 2406
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"

ORIGIN
Query Match 100.0%; Score 2388; DB 6; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTA 60
Db 18 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTA 77
QY 61 TATAGATGGAACACAGCGACGCAAAAACCGGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db 78 TATAGATGGAACACAGCGACGCAAAAACCGGAGAAATTTGACTCCTGATGAGGTTAGCAA 137
QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATTAACAGACCAAGGCTATGTCAC 180
Db 138 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATTAACAGACCAAGGCTATGTCAC 197
QY 181 TTCACATGGCGACCACTATCATTTATACATGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db 198 TTCACATGGCGACCACTATCATTTATACATGTAAGGTTCCCTTATGACGCTATCATCAG 257
QY 241 TGAAGATTTACTCATGAAGATCCAACTATAGCTAAGCTAAGAGATGAGGATTTGTTATGA 300
Db 258 TGAAGATTTACTCATGAAGATCCAACTATAGCTAAGCTAAGAGATGAGGATTTGTTATGA 317
QY 301 GGTCAAGGGTGGATATGTTATCAAGCTAGATGAAATATACTATGTTTACCTTAAGATGC 360
Db 318 GGTCAAGGGTGGATATGTTATCAAGCTAGATGAAATATACTATGTTTACCTTAAGATGC 377
QY 361 TGCCCGCGGGATAACGTCGTAACAAAGAGGAAATCAATCGCAAAAACAGAGCATAG 420
Db 378 TGCCCGCGGGATAACGTCGTAACAAAGAGGAAATCAATCGCAAAAACAGAGCATAG 437
QY 421 TCAACATCGTGAAGGTGAACCTCAAGAAACGATGGTGTGCTTGGCTTGCACGTTGCGCA 480
```

1561 AGCTGATAGTATCAACGTCAGATGGTTTACATTTTGTGATGAACATGATATTAATCAGTGA 1620
 1578 AGCTGATAGTATCAACGTCAGATGGTTTACATTTTGTGATGAACATGATATTAATCAGTGA 1637
 1621 TGAAGGAGATGATATGTAACGCTCATATGGCCCATAGTCACTGATTTGGAAGAGATAG 1680
 1638 TGAAGGAGATGATATGTAACGCTCATATGGCCCATAGTCACTGATTTGGAAGAGATAG 1697
 1681 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCCTATATCTAAAGAAAAAGGTTATCCTACC 1740
 1698 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCCTATATCTAAAGAAAAAGGTTATCCTACC 1757
 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1800
 1758 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1817
 1801 CAATCGTGTGAAGGGGAAAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1860
 1818 CAATCGTGTGAAGGGGAAAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1877
 1861 TACAGTTGAGTTAAAGCGGTAATTTGATTTTCTCTATTAAGGATCATTAACATAATAT 1920
 1878 TACAGTTGAGTTAAAGCGGTAATTTGATTTTCTCTATTAAGGATCATTAACATAATAT 1937
 1921 TAAATTTGCTTGGTTTGGATGATCACATACAAAGCTCCAAATGCTATACCTTGGGAAGA 1980
 1938 TAAATTTGCTTGGTTTGGATGATCACATACAAAGCTCCAAATGCTATACCTTGGGAAGA 1997
 1981 TTTGTTGCGAGTAATAGTACTAGTAGAAACACCTCTGACGAACGTCACATCTTAATGA 2040
 1998 TTTGTTGCGAGTAATAGTACTAGTAGAAACACCTCTGACGAACGTCACATCTTAATGA 2057
 2041 TGGATGGGCAATGCCAGTACGATGTTAGGCAAGAAAGCAAGCAAGTGAAGATCCAAA 2100
 2058 TGGATGGGCAATGCCAGTACGATGTTAGGCAAGAAAGCAAGTGAAGATCCAAA 2117
 2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTACGAGAAACACCTCTGAGCCAGAAAGTCCC 2160
 2118 TAAGAACTTCAAGCGGATGAAGAGCCAGTACGAGAAACACCTCTGAGCCAGAAAGTCCC 2177
 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGCAA 2220
 2178 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGCAA 2237
 2221 AGTAACGATTTCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
 2238 AGTAACGATTTCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2297
 2281 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAAAGCAAGAAATTAATCTTC 2340
 2298 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAAAGCAAGAAATTAATCTTC 2357
 2341 GTTGTAAAGGAAAGTAATCCTTCTATCTGTAGTAAGGAAAAAATAAAC 2389
 2358 GTTGTAAAGGAAAGTAATCCTTCTATCTGTAGTAAGGAAAAAATAAAC 2406

RESULT 5
 BD268048
 LOCUS
 DEFINITION Streptococcus pneumoniae protein and immunogenic fragments for vaccines.
 ACCESSION BD268048
 VERSION BD268048.1 GI:33077816
 KEYWORDS JP 2002532561-A/6.
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1. (bases 1 to 2451)
 Johnson, L.S., Koenig, S. and Adamou, J.E.
 Streptococcus pneumoniae protein and immunogenic fragments for

JOURNAL Patent JP 2002532561-A 6 02-OCT-2002;
 MEDIMUNE INC
 COMMENT OS Streptococcus pneumoniae
 PN JP 2002532561-A/6
 PD 02-OCT-2002
 PF 21-DEC-1999 JP 2000589215
 PR 21-DEC-1998 US 60/113048
 PI LESLIE S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU
 PC A61K39/09, A61K38/00, A61P31/10, A61K37/02
 CC Streptococcus pneumoniae protein and immunogenic fragments for
 vaccines
 PH Key Location/Qualifiers
 FT source 1..2451
 FT /organism='Streptococcus pneumoniae'
 /mol_type='genomic DNA'
 /db_xref='taxon:1313'
 FEATURES
 source
 ORIGIN
 Query Match 100.0%; Score 2388; DB 6; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTTACGAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
 Db 60 TTTTACGAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 119
 QY 61 TATAGATGAAAAAACAAGCGAGCGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 120
 Db 120 TATAGATGAAAAAACAAGCGAGCGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 179
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCCAC 180
 Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCCAC 239
 QY 181 TTCATGCGGACCACTATCAATTTATCAATGTTAAGGTTCTTTATGACGCTATCATCAG 240
 Db 240 TTCATGCGGACCACTATCAATTTATCAATGTTAAGGTTCTTTATGACGCTATCATCAG 299
 QY 241 TGAAGAAATTTCTATGAAGATCCAACTATAAGCTAAAGTAAAGTATGTTAAATCA 300
 Db 300 TGAAGAAATTTCTATGAAGATCCAACTATAAGCTAAAGTAAAGTATGTTAAATCA 359
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC 360
 Db 360 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC 419
 QY 361 TGCCACGCGGATACGTTCCGTACAAAGAGAGAAATCAATCGAAGAAACAGAGCATAG 420
 Db 420 TGCCACGCGGATACGTTCCGTACAAAGAGAGAGAAATCAATCGAAGAAACAGAGCATAG 479
 QY 421 TCAACATCGTGAAGGTGGAACCTCAAGAAAGAGTGGTCTGTTGCTTGGCAACGTTCCCA 480
 Db 480 TCAACATCGTGAAGGTGGAACCTCAAGAAAGAGTGGTCTGTTGCTTGGCAACGTTCCCA 539
 QY 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTTCTGATATCATAGAGGATAC 540
 Db 540 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTTCTGATATCATAGAGGATAC 599
 QY 541 TGGTATGCTTATATCTGTTCTCTCATGAGATCATTAACATTAATCTCTAAGATGAGTT 600
 Db 600 TGGTATGCTTATATCTGTTCTCTCATGAGATCATTAACATTAATCTCTAAGATGAGTT 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAGCGCTTCTATCTGCTGAGGAAATCTGTCAA 660
 Db 660 ATCAGCTAGCGAGTTGGCTGCTGAGAGCGCTTCTATCTGCTGAGGAAATCTGTCAA 719
 QY 661 TTCAAGAACCTTATCGCCGACAAAAATAGGATAAACACTTCAAGAAACAACTGGTACCTTC 720
 Db 720 TTCAAGAACCTTATCGCCGACAAAAATAGGATAAACACTTCAAGAAACAACTGGTACCTTC 779

QY 721 TGTAAAGCAATCCAGGAATACAAATATACTAAACAAGCAACAAAGCAAGCACTAAACAGTCA 780
 Db 780 TGTAAAGCAATCCAGGAATACAAATATACTAAACAAGCAACAAAGCAAGCACTAAACAGTCA 839
 QY 781 AGCAAGTCAAAGTAATGACATGATGATGCTCTTTGAAACAGCTCTACAAACTGCTTTGAG 840
 Db 840 AGCAAGTCAAAGTAATGACATGATGATGCTCTTTGAAACAGCTCTACAAACTGCTTTGAG 899
 QY 841 TCAACGACATGTAGAACTGATGCGCTTGTCTTTGATCCAGCAAAATCAACAGTCAAC 900
 Db 900 TCAACGACATGTAGAACTGATGCGCTTGTCTTTGATCCAGCAAAATCAACAGTCAAC 959
 QY 901 AGCTAGAGGTGTTGCGAGTGCACACGAGAGATCAATACCACTTCCCTTACTCTCAAAAT 960
 Db 960 AGCTAGAGGTGTTGCGAGTGCACACGAGAGATCAATACCACTTCCCTTACTCTCAAAAT 1019
 QY 961 GTCTGAATTTGGAAGAACGAATCGCTGATATATTCCTCCCTTCTGTTATGTTCAAACTATG 1020
 Db 1020 GTCTGAATTTGGAAGAACGAATCGCTGATATATTCCTCCCTTCTGTTATGTTCAAACTATG 1079
 QY 1021 GGTACAGATTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
 Db 1080 GGTACAGATTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1139
 QY 1081 CCGGCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTTTGTTAGTCAAGCTGGT 1140
 Db 1140 CCGGCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTTTGTTAGTCAAGCTGGT 1199
 QY 1141 ACGAAAGTTGGGGAAGGATATGATTCGAAGAAAGGGGATCTCTGTTATGTTCTTTGTC 1200
 Db 1200 ACGAAAGTTGGGGAAGGATATGATTCGAAGAAAGGGGATCTCTGTTATGTTCTTTGTC 1259
 QY 1201 GAAAGATTTTACCCTGAACTGTAATTTGAAAGCAAGTATCAAAACAGAGAG 1260
 Db 1260 GAAAGATTTTACCCTGAACTGTAATTTGAAAGCAAGTATCAAAACAGAGAG 1319
 QY 1261 TGTTTTCACACATTTAACTGCTAAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 1320
 Db 1320 TGTTTTCACACATTTAACTGCTAAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 1379
 QY 1321 TGATAAGCATATAATCTGTTAACTAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG 1380
 Db 1380 TGATAAGCATATAATCTGTTAACTAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG 1439
 QY 1381 TAATTCGTATTTCCAGCCCTTAGACAAATTAATAGAACGCTTCAATGATCAATCGACTAA 1440
 Db 1440 TAATTCGTATTTCCAGCCCTTAGACAAATTAATAGAACGCTTCAATGATCAATCGACTAA 1499
 QY 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCAACCAATTAACCCATCCAGAGC 1500
 Db 1500 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCAACCAATTAACCCATCCAGAGC 1559
 QY 1501 ACTTGGCAACCAAAATCTCAAAATGAGTATCTAGAGCAAGTTCGATTTCTCAAT 1560
 Db 1560 ACTTGGCAACCAAAATCTCAAAATGAGTATCTAGAGCAAGTTCGATTTCTCAAT 1619
 QY 1561 AGCTGATAAGTATACAAAGCTGATGTTTACATTTTGTGAAACATGATATCAATCAATG 1620
 Db 1620 AGCTGATAAGTATACAAAGCTGATGTTTACATTTTGTGAAACATGATATCAATCAATG 1679
 QY 1621 TGAAGGAGATGCTATATGTAACGCTCATATGCGCCATAGTCACTGGATTTGAAAGATAG 1680
 Db 1680 TGAAGGAGATGCTATATGTAACGCTCATATGCGCCATAGTCACTGGATTTGAAAGATAG 1739
 QY 1681 CCTTTCTGATAAGGAAAGGTTGAGCTCAAGCTTATCTTAAAGAAAGGATCTCTTACC 1740
 Db 1740 CCTTTCTGATAAGGAAAGGTTGAGCTCAAGCTTATCTTAAAGAAAGGATCTCTTACC 1799
 QY 1741 TCCATCTCCAGACGAGATGTTAAGCAAAATCCAACCTGGAGATGATGACAGCTATTTTA 1800
 Db 1800 TCCATCTCCAGACGAGATGTTAAGCAAAATCCAACCTGGAGATGATGACAGCTATTTTA 1859

QY 1801 CAATCTGTGAAAGGGGAAAAACGAATTCACATCTGTTGCACTTCCATATATGTTGAGCA 1860
 Db 1860 CAATCTGTGAAAGGGGAAAAACGAATTCACATCTGTTGCACTTCCATATATGTTGAGCA 1919
 QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTAATTCCTCATAGGATCAATACCATATAT 1920
 Db 1920 TACAGTTGAGGTTAAAAACGGTAATTTGATTAATTCCTCATAGGATCAATACCATATAT 1979
 QY 1921 TAAATTTGCTGTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
 Db 1980 TAAATTTGCTGTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 2039
 QY 1981 TTTGTTTGGACAGATTAAAGTACTACGTAGAACCCCTGACGAACGCTCCACATTTCTATGA 2040
 Db 2040 TTTGTTTGGACAGATTAAAGTACTACGTAGAACCCCTGACGAACGCTCCACATTTCTATGA 2099
 QY 2041 TGGATGGGCAATTCGCAGTGAAGTATGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2100
 Db 2100 TGGATGGGCAATTCGCAGTGAAGTATGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2159
 QY 2101 TTAAGAACTTCAAAGCCGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAAGTCCC 2160
 Db 2160 TTAAGAACTTCAAAGCCGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAAGTCCC 2219
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGACAGAAAGTTTGTTCGAA 2220
 Db 2220 TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGACAGAAAGTTTGTTCGAA 2279
 QY 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA 2280
 Db 2280 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA 2339
 QY 2281 TTTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAAGACAGAAAAATTTACTTGC 2340
 Db 2340 TTTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAAGACAGAAAAATTTACTTGC 2399
 QY 2341 GTTGTAAAGGAGTAATCTCTTCACTGCTGTAAGTAAGGAAAAATAAAC 2389
 Db 2400 GTTGTAAAGGAGTAATCTCTTCACTGCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 6
 AR344447
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source

AR344447
 Sequence 9 from patent US:6582706.
 AR344447
 AR344447.1 GI:33740486
 Unknown.
 Unclassified.
 Johnson, L.S. and Adamou, J.E.
 Vaccine compositions comprising Streptococcus pneumoniae
 polypeptides having selected structural MOTIFS
 Patent: US 6582706-A 9 24-JUN-2003;
 Location/Qualifiers
 1..2451
 /organism="unknown"
 /mol_type="genomic DNA"



Query Match 100.0%; Score 2388; DB 6; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTTAAAGGAAAAATTCGTTCTTCTTA 60
 Db 60 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTTAAAGGAAAAATTCGTTCTTCTTA 119
 QY 61 TATAGATGGAACCAACAGCGACGCAAAACCGGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 Db 120 TATAGATGGAACCAACAGCGACGCAAAACCGGAGAAATTTGACTCTCTGATGAGTTAGCAA 179

QY	121	GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGACAAGGCTATGTCCAC	180
DB	180	GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGACAAGGCTATGTCCAC	239
QY	181	TTCCACATGGCGGACCACTATCATTTATTACANTGTAAGGTTCCCTTATGACGCTATCATCAG	240
DB	240	TTCCACATGGCGGACCACTATCATTTATTACANTGTAAGGTTCCCTTATGACGCTATCATCAG	299
QY	241	TGAAGAAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA	300
DB	300	TGAAGAAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA	359
QY	301	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAAAATACATATGTTTACCTTAAGGATGC	360
DB	360	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAAAATACATATGTTTACCTTAAGGATGC	419
QY	361	TGCCCCGCGGATAAAGCTCGGTACAAAGAGGAAAAACAATCGACAAAAACAAGAGCATAG	420
DB	420	TGCCCCGCGGATAAAGCTCGGTACAAAGAGGAAAAACAATCGACAAAAACAAGAGCATAG	479
QY	421	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGCTTGCCTTGGCACGTTTCGCA	480
DB	480	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGCTTGCCTTGGCACGTTTCGCA	539
QY	481	AGGACGCTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	540
DB	540	AGGACGCTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	599
QY	541	TGTTGATGCTTTATCGTTTCTCATGAGATCATTTACATTTACATTCCTTAAGAAATCAGTT	600
DB	600	TGTTGATGCTTTATCGTTTCTCATGAGATCATTTACATTTACATTCCTTAAGAAATCAGTT	659
QY	601	ATCAGCTAGCGAGTTGGCTGCTCGAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA	660
DB	660	ATCAGCTAGCGAGTTGGCTGCTCGAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAA	719
QY	661	TTCAAGAACCTATCGCCGACAAATAGCGATACACTTCAAGACAAACTGGGTACCTTTC	720
DB	720	TTCAAGAACCTATCGCCGACAAATAGCGATACACTTCAAGACAAACTGGGTACCTTTC	779
QY	721	TGTAAGCAATCCAGGAACACTCAATACTAACAAGCAACAACAGCAAACTAACAGTCA	780
DB	780	TGTAAGCAATCCAGGAACACTCAATACTAACAAGCAACAACAGCAAACTAACAGTCA	839
QY	781	AGCAAGTCAAGTAATGACATTGATAGTCTCTTTGAAAACAGCTCTACAAACTGGCTTTGAG	840
DB	840	AGCAAGTCAAGTAATGACATTGATAGTCTCTTTGAAAACAGCTCTACAAACTGGCTTTGAG	899
QY	841	TCAACGACATGTAGAACTCTGATGCCCTTGCTTTGATCCAGCAAAATCAACAGTCCGAC	900
DB	900	TCAACGACATGTAGAACTCTGATGCCCTTGCTTTGATCCAGCAAAATCAACAGTCCGAC	959
QY	901	AGCTAGAGGTGTTGCAGTGGCCACACGAGATCAATTACACTTCATCCCTTACTCTCAAAAT	960
DB	960	AGCTAGAGGTGTTGCAGTGGCCACACGAGATCAATTACACTTCATCCCTTACTCTCAAAAT	1019
QY	961	GTCTGAATTCGAAGAACGAATCGCTCGTATTATTCCTCCCTTCGTTATCGTTTCAAAACATTG	1020
DB	1020	GTCTGAATTCGAAGAACGAATCGCTCGTATTATTCCTCCCTTCGTTATCGTTTCAAAACATTG	1079
QY	1021	GGTACCAGATTCAAGGCCAGAACCAAGTCCACAAACCGACTCCGAAACCTAGTCCAGG	1080
DB	1080	GGTACCAGATTCAAGGCCAGAACCAAGTCCACAAACCGACTCCGAAACCTAGTCCAGG	1139
QY	1081	CCCGCAACCTGCACAAATCTTAAATAGACTCAAATTTCTTTGGTTAGTCAGCTGGT	1140
DB	1140	CCCGCAACCTGCACAAATCTTAAATAGACTCAAATTTCTTTGGTTAGTCAGCTGGT	1199
QY	1141	ACGNAAGTTGGGGAAGGATATGTTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTTC	1200
DB	1200	ACGNAAGTTGGGGAAGGATATGTTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTTC	1259

QY	1201	GAAGAATTTACCATCTGAAACTGTTAAAAATCTTTGAAGCAAGATTATCAAAAACAAGAGAG	1261
DB	1260	GAAGAATTTACCATCTGAAACTGTTAAAAATCTTTGAAGCAAGATTATCAAAAACAAGAGAG	1319
QY	1261	TGTTTCCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCACGAATTTTAA	1320
DB	1320	TGTTTCCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCACGAATTTTAA	1379
QY	1321	TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTTGNAAAATAAGGCTCG	1380
DB	1380	TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTTGNAAAATAAGGCTCG	1439
QY	1381	TAATTCGTGATTTCCAGCCCTTAGACAAATATTATGAACCGCTTGAATGATGAATCGACTAA	1440
DB	1440	TAATTCGTGATTTCCAGCCCTTAGACAAATATTATGAACCGCTTGAATGATGAATCGACTAA	1499
QY	1441	TAAAGAAAAATTCGTAGATGATATTATTTGGCATTCCTAGACCAATATACCCATCCAGAGCG	1500
DB	1500	TAAAGAAAAATTCGTAGATGATATTATTTGGCATTCCTAGACCAATATACCCATCCAGAGCG	1559
QY	1501	ACTTTGGCAAAACCAAATTCCTCAAAATTGAGTATACCTGAAGACGAAGTTCGTAATTCCTCAATT	1560
DB	1560	ACTTTGGCAAAACCAAATTCCTCAAAATTGAGTATACCTGAAGACGAAGTTCGTAATTCCTCAATT	1619
QY	1561	AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA	1620
DB	1620	AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA	1679
QY	1621	TGAAGGAGATGCATATGCTAAGCCTCATATGGCCCATAGTCACTGGATTCGGAAGAAGTAG	1680
DB	1680	TGAAGGAGATGCATATGCTAAGCCTCATATGGCCCATAGTCACTGGATTCGGAAGAAGTAG	1739
QY	1681	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTTACTTAAAGAAAAAGGTATCCTTACC	1740
DB	1740	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTTACTTAAAGAAAAAGGTATCCTTACC	1799
QY	1741	TCCATCTCCAGACGCAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTTA	1800
DB	1800	TCCATCTCCAGACGCAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTTA	1859
QY	1801	CAATCGTGTGAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGTTGAGCA	1860
DB	1860	CAATCGTGTGAAGGGGAAAAACGAATTCCACTCGTTCGACTTCCATATATGTTGAGCA	1919
QY	1861	TACAGTTTGAGGTTAAAAACGGTAATTTGATTATTTCCTCATAAAGATCATTTACCATAATAT	1920
DB	1920	TACAGTTTGAGGTTAAAAACGGTAATTTGATTATTTCCTCATAAAGATCATTTACCATAATAT	1979
QY	1921	TAAATTTGCTTGGTTTGATGATCACACATACAAAGTCCAAATGGCTTATACCTTGGGAAGA	1980
DB	1980	TAAATTTGCTTGGTTTGATGATCACACATACAAAGTCCAAATGGCTTATACCTTGGGAAGA	2039
QY	1981	TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA	2040
DB	2040	TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA	2099
QY	2041	TGGATGGGCAATGCCAGTCAGCATGTGTTAGGCAAGAAAAAGACCAAGTCAAGATCCAAA	2100
DB	2100	TGGATGGGCAATGCCAGTCAGCATGTGTTAGGCAAGAAAAAGACCAAGTCAAGATCCAAA	2159
QY	2101	TAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC	2160
DB	2160	TAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC	2219
QY	2161	TCAAGTAGAGACTGMAAAAGTTAGAGCCCAACTCAAGAGACGAGAGTTTGTCTTCGGAA	2220
DB	2220	TCAAGTAGAGACTGMAAAAGTTAGAGCCCAACTCAAGAGACGAGAGTTTGTCTTCGGAA	2279
QY	2221	AGTAAACGGATTCAGTCTGAAAGCCCAATGCAACAGAAACCTCTAGCTGTTTACGAATAA	2280
DB	2280	AGTAAACGGATTCAGTCTGAAAGCCCAATGCAACAGAAACCTCTAGCTGTTTACGAATAA	2339
QY	2281	TTTGACTCTTTCAAATATATGATATAACATAGTATCATCGGCAAGACGAGAAAAATTACTTGC	2340


```
Db      1230  CCGCAACCTGACCAAACTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT 1289
QY      1141  AGGAAAGTTGGGGAAGATATGTAATTCGAAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1200
Db      1290  ACGAAAGTTGGGGAAGATATGTAATTCGAAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1349
QY      1201  GAAAGATTACCATCTGAAACCTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
Db      1350  GAAAGATTACCATCTGAAACCTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAGAG 1409
QY      1261  TGTTCACACACTTTAACTGCTAAAAAGAAATGTGCTCTCGTGACCAAGAAATTTTA 1320
Db      1410  TGTTCACACACTTTAACTGCTAAAAAGAAATGTGCTCTCGTGACCAAGAAATTTTA 1469
QY      1321  TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTGNAATAAGGGTGC 1380
Db      1470  TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTGNAATAAGGGTGC 1529
QY      1381  TAAATTCGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db      1530  TAAATTCGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1589
QY      1441  TAAAGAAATTTGGTAGATGATTTATTTGGCATTCCTPAGCACCAATTAACCATCCAGAGCG 1500
Db      1590  TAAAGAAATTTGGTAGATGATTTATTTGGCATTCCTPAGCACCAATTAACCATCCAGAGCG 1649
QY      1501  ACTTGGCAACCAAAATCTCAATTCAGTATGTAAGAGCAAGTTTCGTTATGCTCAATT 1560
Db      1650  ACTTGGCAACCAAAATCTCAATTCAGTATGTAAGAGCAAGTTTCGTTATGCTCAATT 1709
QY      1561  AGCTGATAGTATACAAACGCTCAGATGGTTACATTTTGTGTAACATGATATATCAGTGA 1620
Db      1710  AGCTGATAGTATACAAACGCTCAGATGGTTACATTTTGTGTAACATGATATATCAGTGA 1769
QY      1621  TGAAGAGATGATATGTAACGCTCATATGCGCCATAGTGGCCATAGTCACTGANTGGAAGATAG 1680
Db      1770  TGAAGAGATGATATGTAACGCTCATATGCGCCATAGTGGCCATAGTCACTGANTGGAAGATAG 1829
QY      1681  CCTTTCTGATAGGAAAGTTGCAGCTCAAGCCTATATCAAGAAAGAGTATCCTACC 1740
Db      1830  CCTTTCTGATAGGAAAGTTGCAGCTCAAGCCTATATCAAGAAAGAGTATCCTACC 1889
QY      1741  TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTTA 1800
Db      1890  TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTTA 1949
QY      1801  CAATCGTGTGAAAGGGGAAAAACGAAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1860
Db      1950  CAATCGTGTGAAAGGGGAAAAACGAAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 2009
QY      1861  TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCATACCATAATAT 1920
Db      2010  TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCATACCATAATAT 2069
QY      1921  TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db      2070  TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 2129
QY      1981  TTTGTTTGGACGATTAAGTACTACTAGAGACACACCTCTGACGACGTCACATTCATGA 2040
Db      2130  TTTGTTTGGACGATTAAGTACTACTAGAGACACACCTCTGACGACGTCACATTCATGA 2189
QY      2041  TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAA 2100
Db      2190  TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAA 2249
QY      2101  TAAGAACTTCAAAGCGGATCAAGAGCCAGTAGAGGAAACCTGCTGAGCCAGAAAGTCCC 2160
Db      2250  TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACCTGCTGAGCCAGAAAGTCCC 2309
QY      2161  TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTGTTCGAA 2220
Db      2310  TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTGTTCGAA 2369
```

```
QY      2221  AGTAACGATTTCTAGTCTGAAGCCCAATGCACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db      2370  AGTAACGATTTCTAGTCTGAAGCCCAATGCACAGAAACTCTAGCTGGTTTACGAAATAA 2429
QY      2281  TTTGACTCTTCAAAATTTATGGATAACAATAGTAGTATCATGCGAGAGCAGAAAAATTTACTGC 2340
Db      2430  TTTGACTCTTCAAAATTTATGGATAACAATAGTAGTATCATGCGAGAGCAGAAAAATTTACTGC 2489
QY      2341  GTTGTAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAT 2389
Db      2490  GTTGTAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAT 2538

RESULT 8
AR218862
LOCUS      AR218862      8195 bp      DNA      linear      PAT 25-SEP-2002
DEFINITION Sequence 94 from patent US 6420135.
ACCESSION AR218862
VERSION   AR218862.1 GI:23319796
KEYWORDS  .
SOURCE    Unknown.
ORGANISM  Unknown.
REFERENCE 1 (bases 1 to 8195)
AUTHORS   Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
          Fannon, M.R., and Dougherty, B.A.
TITLE      Streptococcus pneumoniae polynucleotides and sequences
JOURNAL    Patent: US 6420135-A 94 16-JUL-2002;
FEATURES   Location/Qualifiers
            source          1..8195
                        /organism="unknown"
                        /mol_type="genomic DNA"

ORIGIN

Query Match      100.0%; Score 2388; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1   TTCCTTAGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA 60
Db      3053  TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 3112
QY      61   TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA 120
Db      3113  TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA 3172
QY      121  GCGTCAAGGAATCAATGTGTAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCA 180
Db      3173  GCGTCAAGGAATCAATGTGTAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCA 3232
QY      181  TTCACATGGCGACCACTATCATTATACATGGTAAAGTTCCCTTATGACGCTATCATCAG 240
Db      3233  TTCACATGGCGACCACTATCATTATACATGGTAAAGTTCCCTTATGACGCTATCATCAG 3292
QY      241  TCAAGAAATTTACTCATGAAGATCCAAACTATAAAGCTAAAAGATGAGGATTTGTTAATCA 300
Db      3293  TGAAGAAATTTACTCATGAAGATCCAAACTATAAAGCTAAAAGATGAGGATTTGTTAATCA 3352
QY      301  GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAAGGATGC 360
Db      3353  GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAAGGATGC 3412
QY      361  TGCCCAACCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 420
Db      3413  TGCCCAACCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 3472
QY      421  TCAACATCGTGAAGTGGAACTCCAAAGAACCGATCGTGTGCTTGTGCTTGGCAGCTTCGCA 480
Db      3473  TCAACATCGTGAAGTGGAACTCCAAAGAACCGATCGTGTGCTTGTGCTTGGCAGCTTCGCA 3532
QY      481  AGGACGCTATACACAGATGATGTTTATATCTTTTAAATGCTTCTGATATCATAGAGTAC 540
```


3533 AGGAGCTATACAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGATAC 3592
541 TGGTGATGCTTATATCGTTCTCTATGAGATCATTTACCATTTACATTTCTTAAGATGATT 600
3593 TGGTGATGCTTATATCGTTCTCTATGAGATCATTTACCATTTACATTTCTTAAGATGATT 3652
601 ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTTATCTGCTGAGAGAAATCTGTCAA 660
3653 ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTTATCTGCTGAGAGAAATCTGTCAA 3712
661 TTCAGAACCTATCCGCGCAAAATAGCGATAACACTTCAAGAACAAATGGGTACCTTC 720
3713 TTCAGAACCTATCCGCGCAAAATAGCGATAACACTTCAAGAACAAATGGGTACCTTC 3772
721 TGTAGCAATCCAGGAATCAAAATCTAACACAGCAACAAACAGCAACACTTAACAGTCA 780
3773 TGTAGCAATCCAGGAATCAAAATCTAACACAGCAACAAACAGCAACACTTAACAGTCA 3832
781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAAACAGCTCTACAAACTGCTTTGAG 840
3833 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAAACAGCTCTACAAACTGCTTTGAG 3892
841 TCACAGCATGTAGAAATCTGAGGCTTGTCTTTGATCCAGCAAAATCAACAGTCAAC 900
3893 TCACAGCATGTAGAAATCTGAGGCTTGTCTTTGATCCAGCAAAATCAACAGTCAAC 3952
901 AGCTAGAGGTTGTCAGTGCACACGAGATCATTTACCACTTCTATCCCTTACTCTCAAT 960
3953 AGCTAGAGGTTGTCAGTGCACACGAGATCATTTACCACTTCTATCCCTTACTCTCAAT 4012
961 GTCTGAATTTGGAAGCAATCGCTCGTATTAATTTCCCTTCTGTTATCGTTCAACACCAT 1020
4013 GTCTGAATTTGGAAGCAATCGCTCGTATTAATTTCCCTTCTGTTATCGTTCAACACCAT 4072
1021 GGTACAGATTCAGGCGCAGAAACAAACAGTCCACACGAGTCCAGTCCAGG 1080
4073 GGTACAGATTCAGGCGCAGAAACAAACAGTCCACACGAGTCCAGTCCAGG 4132
1081 CCGCAACCTGACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 1140
4133 CCGCAACCTGACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 4192
1141 ACAGAAAGTTGGGAAGGATATGTAATTCGAAGAAAGGCGATCTCTGTTATGCTTTTC 1200
4193 ACAGAAAGTTGGGAAGGATATGTAATTCGAAGAAAGGCGATCTCTGTTATGCTTTTC 4252
1201 GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1260
4253 GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 4312
1261 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCACCAAGAAATTTTA 1320
4313 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCACCAAGAAATTTTA 4372
1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATPAAAGCCTTGTGTTGAAATTAAGGCTG 1380
4373 TGATAAGCATATAATCTGTTAACTGAGGCTCATPAAAGCCTTGTGTTGAAATTAAGGCTG 4432
1381 TAAATCTGATTTCCAGCCTTAGCAATTTATAGACGCTTGAATGATGATGATGATGATG 1440
4433 TAAATCTGATTTCCAGCCTTAGCAATTTATAGACGCTTGAATGATGATGATGATGATGAT 4492
1441 TAAAGAAATTTGTTAGATGATTTATTTGGCAATCTCTAGCAACCAATACCCATCCAGAGG 1500
4493 TAAAGAAATTTGTTAGATGATTTATTTGGCAATCTCTAGCAACCAATACCCATCCAGAGG 4552
1501 ACTTGGCAACCAATTTCTAAATGAGTATCTGAGAGCAAGTTCGATTTGCTCAATT 1560
4553 ACTTGGCAACCAATTTCTAAATGAGTATCTGAGAGCAAGTTCGATTTGCTCAATT 4612
1561 AGCTGATAAGTATCAACGTCAGATGTTTACATTTTGTGCAACATCATATATCACTGA 1620
4613 AGCTGATAAGTATCAACGTCAGATGTTTACATTTTGTGCAACATCATATATCACTGA 4672

Qy 1621 TGAAGGAGATGCATATCTTAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
Db 4673 TGAAGGAGATGCATATCTTAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 4732
Qy 1681 CCTTCTGTATAAGGAAAGTTGAGCTCAAGCCTATCTACTAAAGAAAGATGATCTTACC 1740
Db 4733 CCTTCTGTATAAGGAAAGTTGAGCTCAAGCCTATCTACTAAAGAAAGATGATCTTACC 4792
Qy 1741 TCAATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGAGCATTTTA 1800
Db 4793 TCAATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGAGCATTTTA 4852
Qy 1801 CAATCGTGTCAAGGGGAAAGCAATTCACCTGTTTCCATCTTCCATATATGTTGAGCA 1860
Db 4853 CAATCGTGTCAAGGGGAAAGCAATTCACCTGTTTCCATCTTCCATATATGTTGAGCA 4912
Qy 1861 TACAGTTGAGGTTTAAAGCGTAAATTTGATTTCTCTCATAGGATCATTTACCAATAT 1920
Db 4913 TACAGTTGAGGTTTAAAGCGTAAATTTGATTTCTCTCATAGGATCATTTACCAATAT 4972
Qy 1921 TAAATTTGCTTGGTTTTCATGATCACACATACAAAGCTCCAAATGGCTTATACCTTGGAGA 1980
Db 4973 TAAATTTGCTTGGTTTTCATGATCACACATACAAAGCTCCAAATGGCTTATACCTTGGAGA 5032
Qy 1981 TTTGTTTGGCACGATTAAGTACTACGTAGAACACCTTGACGACCTCCACATTTCTAATGA 2040
Db 5033 TTTGTTTGGCACGATTAAGTACTACGTAGAACACCTTGACGACCTCCACATTTCTAATGA 5092
Qy 2041 TGGATGGGCAATGCCAGTGAGCATGTTTGGCAAGAAAGACCAAGTGAAGATCCAA 2100
Db 5093 TGGATGGGCAATGCCAGTGAGCATGTTTGGCAAGAAAGACCAAGTGAAGATCCAA 5152
Qy 2101 TAAGAACTTCAAGCGGATCAAGAGCCAGTAGAGGAAACCTCTGCTGAGCGAGAGTCCC 2160
Db 5153 TAAGAACTTCAAGCGGATCAAGAGCCAGTAGAGGAAACCTCTGCTGAGCGAGAGTCCC 5212
Qy 2161 TCAAGTAGACACTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTTCGGAA 2220
Db 5213 TCAAGTAGACACTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTTCGGAA 5272
Qy 2221 AGTAACGGATTCAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA 2280
Db 5273 AGTAACGGATTCAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA 5332
Qy 2281 TTTGACTCTTCAAAATTTGATACATAGTATCTATGCGAGAGCAGAAATTTACTTGC 2340
Db 5333 TTTGACTCTTCAAAATTTGATACATAGTATCTATGCGAGAGCAGAAATTTACTTGC 5392
Qy 2341 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 5393 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 9

BD003774
LOCUS BD003774 8195 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003774
VERSION BD003774.1 GI:18631735
KEYWORDS JP 2001501833-A/94.
SOURCE unidentified
ORGANISM Streptococcus pneumoniae
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
Fannon, M., and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/94
FD 13-FEB-2001

PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..8195 /organism='Unidentified'.
FT Location/Qualifiers
1..8195
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

FEATURES
source

ORIGIN

Query Match 100.0%; Score 2388; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
Db 3053 TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 3112

QY 61 TATAGATGAAACAAAGCGACGCAAAAAACGGAGAATTTGACTCTCTGATGAGGTAGCAA 120
Db 3113 TATAGATGAAACAAAGCGACGCAAAAAACGGAGAATTTGACTCTCTGATGAGGTAGCAA 3172

QY 121 GGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 3173 GGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232

QY 181 TTCACATGGCGACCATCTATTTACAAATGCTAGAGTTTCTTATGACGCTATCATCAG 240
Db 3233 TTCACATGGCGACCATCTATTTACAAATGCTAGAGTTTCTTATGACGCTATCATCAG 3292

QY 241 TGAAGAAATTTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATATTGTTAATGA 300
Db 3293 TGAAGAAATTTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATATTGTTAATGA 3352

QY 301 GGTCAAGGTTGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC 360
Db 3353 GGTCAAGGTTGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC 3412

QY 361 TGCCACGCGGATAGCTCGGTACAAAGAGGGAATCAATCGCAAAACAAAGAGCATAG 420
Db 3413 TGCCACGCGGATAGCTCGGTACAAAGAGGGAATCAATCGCAAAACAAAGAGCATAG 3472

QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAACGATGGTGTGTTGCTTGGCAGCTTCGCA 480
Db 3473 TCAACATCGTGAAGTGGAACTCCAAAGAACGATGGTGTGTTGCTTGGCAGCTTCGCA 3532

QY 481 AGAAGCTATATCAAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
Db 3533 AGAAGCTATATCAAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592

QY 541 TGGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATCATCTTCTTAAGAAATGAGTT 600
Db 3593 TGGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATCATCTTCTTAAGAAATGAGTT 3652

QY 601 ATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCCTTATCTGTCGAGGAAATCTGTCAAA 660
Db 3653 ATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCCTTATCTGTCGAGGAAATCTGTCAAA 3712

QY 661 TTCAGAAACCTATCGCGCAAAATAGCGATAACATTCAGAAACAAACCTGGGATACCTTC 720
Db 3713 TTCAGAAACCTATCGCGCAAAATAGCGATAACATTCAGAAACAAACCTGGGATACCTTC 3772

QY 721 TGTAGCAATCCAGNACTTACAAATACTAACAACACCAACACAGCAACACTAACAGTCA 780
Db 3773 TGTAAAGCAATCCAGGAACCTACAAATACTAACAACCAACAGCAACACTAACAGTCA 3832

QY 781 AGCAAGTCAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAAACTGCCTTTGAG 840
Db 3833 AGCAAGTCAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAAACTGCCTTTGAG 3892

QY 841 TCAACGACATGTAGAAATCTGATGGGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAC 900
Db 3893 TCAACGACATGTAGAAATCTGATGGGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAC 3952

QY 901 AGCTAGAGGTGTCAGTCCACACGAGATCATTTACCACTTCACTCCCTTACTCTCAAT 960
Db 3953 AGCTAGAGGTGTCAGTCCACACGAGATCATTTACCACTTCACTCCCTTACTCTCAAT 4012

QY 961 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTTCGTTATCGTTTCAAAACCATGG 1020
Db 4013 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTTCGTTATCGTTTCAAAACCATGG 4072

QY 1021 GGTACAGATTTCAAGCCAGAAACAAAGTCCAAACCGAGCTCCGGAACCTAGTCCAGG 1080
Db 4073 GGTACAGATTTCAAGCCAGAAACAAAGTCCAAACCGAGCTCCGGAACCTAGTCCAGG 4132

QY 1081 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAAGTGT 1140
Db 4133 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAAGTGT 4192

QY 1141 ACGAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGC 1200
Db 4193 ACGAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGC 4252

QY 1201 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
Db 4253 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 4312

QY 1261 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 1320
Db 4313 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 4372

QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 1380
Db 4373 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 4432

QY 1381 TAAATCTGATTTCCAAAGCTTAGACAAATATATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db 4433 TAAATCTGATTTCCAAAGCTTAGACAAATATATTAGAACGCTTGAATGATGAATCGACTAA 4492

QY 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCATTCTTAGCACCATAATACCCTCCAGAGCG 1500
Db 4493 TAAAGAAATTTGGTAGATGATTTATTTGGCATTCTTAGCACCATAATACCCTCCAGAGCG 4552

QY 1501 ACTTGGCAACCAAAATTTCTCAAATTTGAGTATACTGAAAGCAAGTTTCGTTATGCTCAAT 1560
Db 4553 ACTTGGCAACCAAAATTTCTCAAATTTGAGTATACTGAAAGCAAGTTTCGTTATGCTCAAT 4612

QY 1561 AGCTGATAGTATACAACTGAGATGTTTACATTTTGTGTAACATGATATATCAAGTGA 1620
Db 4613 AGCTGATAGTATACAACTGAGATGTTTACATTTTGTGTAACATGATATATCAAGTGA 4672

QY 1621 TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 1680
Db 4673 TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 4732

QY 1681 CCTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCTTATCTAAAGAAAAAGGATCTCTACC 1740
Db 4733 CCTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCTTATCTAAAGAAAAAGGATCTCTACC 4792

QY 1741 TCCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGAGATAGTGGAGAGCTATTATTA 1800
Db 4793 TCCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGAGATAGTGGAGAGCTATTATTA 4852

QY 1801 CAATCGTGTGAAGGGGAAAAACGAATTCCTCGTTCGACTTCCATATATGTTGAGCA 1860

Db	4853	CAATCGTGTGAAGGGGAAACGAATCCACTGTTCCGACTTCCATATATGTTGACCA	4912	
Qy	1861	TACAGTTGAGGTTAAAACGGTAATTTGATTAATTCCTCATAGGATCATATACCAATATAT	1920	
Db	4913	TACAGTTGAGGTTAAAACGGTAATTTGATTAATTCCTCATAGGATCATATACCAATATAT	4972	
Qy	1921	TAAATTTGCTTGTGTTTGTATGATACACATACAAAGCTCCAAATGCTATATCTTTGGAAGA	1980	
Db	4973	TAAATTTGCTTGTGTTTGTATGATACACATACAAAGCTCCAAATGCTATATCTTTGGAAGA	5032	
Qy	1981	TTTCTTTTGCAGCAATTAAGTACTACGTAGAACACCCCTGACGAACGTCACATCTTAATGA	2040	
Db	5033	TTTCTTTTGCAGCAATTAAGTACTACGTAGAACACCCCTGACGAACGTCACATCTTAATGA	5092	
Qy	2041	TGGATGGGGCNATCCAGTACGATGTTAGGCAAGAACACACACAGTGAAGATCCAAA	2100	
Db	5093	TGGATGGGGCNATCCAGTACGATGTTAGGCAAGAACACACACAGTGAAGATCCAAA	5152	
Qy	2101	TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAACTCC	2160	
Db	5153	TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAACTCC	5212	
Qy	2161	TCAAGTAGAGACTGAAAAGTGAAGGCCAACTCAAAGAGCAGAACTTTTGTTCGCA	2220	
Db	5213	TCAAGTAGAGACTGAAAAGTGAAGGCCAACTCAAAGAGCAGAACTTTTGTTCGCA	5272	
Qy	2221	AGTAACGGATCTAGTCTGAAGCCATGCAACGAACTCTAGCTGTTTACGAATAA	2280	
Db	5273	AGTAACGGATCTAGTCTGAAGCCATGCAACGAACTCTAGCTGTTTACGAATAA	5332	
Qy	2281	TTTGACTCTTCAAATTTGGATATCAATAGTATCATGCGAAGCAGAAATAATTAAC	2340	
Db	5333	TTTGACTCTTCAAATTTGGATATCAATAGTATCATGCGAAGCAGAAATAATTAAC	5392	
Qy	2341	GTGTTTAAAGGAAGTAACTCTTCATCTGTAAGTAAGCAAAAAATAAC	2389	
Db	5393	GTGTTTAAAGGAAGTAACTCTTCATCTGTAAGTAAGCAAAAAATAAC	5441	
RESULT 10				
AE007418/c		10256 bp	DNA	linear
LOCUS	AE007418	10256 bp	DNA	linear
DEFINITION	Streptococcus pneumoniae TIGR4	section 101 of 194	BCT 31-AUG-2001	
ACCESSION	AE007418	AE005672		
VERSION	AE007418.1	GI:14972649		
KEYWORDS	Streptococcus pneumoniae TIGR4			
SOURCE	Streptococcus pneumoniae TIGR4			
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
REFERENCE	1 (bases 1 to 10256)			
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.			
TITLE	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae			
JOURNAL	Science	293 (5529), 498-506	(2001)	
MEDLINE	21357209			
PUBMED	11463916			
REFERENCE	2 (bases 1 to 10256)			
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A.,			
Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.				
Direct Submission				
Submitted (29-JUN-2001)	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
Location/Qualifiers				
1. 10256				
/organism="Streptococcus pneumoniae TIGR4"				
/mol_type="genomic DNA"				
/strain="TIGR4"				
/db_xref="taxon:170187"				
complement(100..702)				
/gene="SP1170"				
complement(100..702)				
/gene="SP1170"				
/notes="identified by Glimmer2; putative"				
/codon_start=1				
/transl_table=11				
/product="hypothetical protein"				
/protein_id="AAK75279.1"				
/db_xref="GI:14972650"				
/translation="MSEVDNEAVNYEFTSDTCLANSIYQSLPKPDKKNFSGDLIP				
TKSPSLVKGDIYGRDSQVNLNRIYFNLYTRKYSLNMRNGCMGDFPHDF				
DIYLDHVAKYAYEQKNNIKYKYLKAILHQENALYFRFFSNDFDELEKNYLTWQ				
VSKETPSEMDFNMEXNISIKIIFERGSKMLNDLKSNNK"				
complement(760..1473)				
/gene="SP1171"				
complement(760..1473)				
/gene="SP1171"				
/notes="identified by match to PFAM protein family HMM PF00702"				
/codon_start=1				
/transl_table=11				
/product="hydrolase, haloacid dehalogenase-like family"				
/protein_id="AAK75280.1"				
/db_xref="GI:14972651"				
/translation="MFYKFLFDLDTLDFDAADVALTQLLKEGVADTQAYKDY				
VPNKALWKDLKTKISQBLVNTFRSLFAHQEKDGSFLAQRYQFYLAQOQSILS				
GAHDLDSLIERDYNLYAATNGITATQGLAQSLGAPYENOVFISLOLQTKEDALF				
YKIQQIAGFSKREKTLMGDSLTADIQGNAGIDITWNPHEHLEHTQAOPTVEY				
SYQDLDDCLDKNILEKITF"				
complement(1724..1843)				
/gene="SP1172"				
complement(1724..1843)				
/gene="SP1172"				
/notes="identified by Glimmer2; putative"				
/codon_start=1				
/transl_table=11				
/product="hypothetical protein"				
/protein_id="AAK75281.1"				
/db_xref="GI:14972652"				
/translation="MLYAVFPFNRSETIYFNCESIKTDCDGAILETFKN"				
complement(1933..2307)				
/gene="SP1173"				
complement(1933..2307)				
/gene="SP1173"				
/notes="identified by Glimmer2; putative"				
/codon_start=1				
/transl_table=11				
/product="conserved hypothetical protein"				
/protein_id="AAK75282.1"				
/db_xref="GI:14972653"				
/translation="MESIGLIVSHSKHIAEGVVELISKVAKDVPIYVRGTEGGGIG				
TSPEQVDRVVSENPAOTLLAFDFLGSAGKMLKMTDFSDKSIINRPIVEGAYNAAA				
LLQAGALSVICQLAELEINK"				
complement(2729..5188)				
/gene="SP1174"				
complement(2729..5188)				
/gene="SP1174"				
/notes="similar to GP:4249624; identified by sequence similarity; putative"				
/codon_start=1				

```

/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75283.1"
/db_xref="GI:14972654"
/translation="MKINKKYLAVLALSVCSYELGRYQAGQKESNRVAYIDG
DOAGQAEMLTPDEVSRRKGINAQIVIKITDQGYVTSFGDHYHYNGKVPYDAIIE
ELLKMDPNYQLKSDSIVNELKGGYVILKNGKYYVLKDAHADNIRTKKEIKRQKEL
SHNHSRADNVAARAQGRYITDDGIIIFNASDIIEDTDGAYIVPHGQHHYIPKNEL
SASELAAYWNGKGSRRSSSYNANPAQRLSENHNLIVTPYHQNGENISSL
LRELYAKPSERHVESDGLIFDPAQITSRARGVAPHGNHYHFIPEYQMSLEKRIA
RIIPRYRNHWVDSRPEPSQPTPEPSPPQAPSPIDEKLVKBAVRKVGQYV
FENGYSYIPAKLISAEATAGIDSLAKQESLHKLTKTDLPSSDREFFYNKAYDL
LARLHODLLNGROWDFEALDNLRELKDVSDKVLVEDILALFAPIRHBERLGP
NAQITVYDEIYQVAKLAGYTTEDGYIFDPRDITSDEGAYVTPMHSHHWIKUSLS
EAERAAQAYAKEGLTTPSTHQSDTEAKGAELYNKRVAAKVPIDRPMYNIQY
TVFVNGSLIIPHYDRHNIKEFWDEGLYAPKTYLLEDLATVRYYYVLEHNRPHS
DNFGNSDHRVKNQKQADNTQTEKPEBKQTEKPEBETPREEKPEQSEKPEPPT
EBEPESPESEPOVETEKVEBKLEAREADLLGKIQDPIIKSNAKETLTGLKNLLPPT
QDNTIMAREKLLALKESK"
complement(5346..7754)
/gene="Sp1175"
CDS
complement(5346..7754)
/gene="Sp1175"
/notes="similar to QP:4249624; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75284.1"
/db_xref="GI:14972655"
/translation="MILSVCSYELGLYQARTVKNRVSVDGKQATOKENTLPDEV
SRREGINARQIVKITDQGYVTSFGDHYHYNGKVPYDAIIEELLKMDPNYKLKDED
IYNEVKGGVIVKDGKYYVLKDAHADNIRTKKEIKRQKESNRREGITPRNDGAV
ALARSQRYITDDGYIFNASDIIEDTDGAYIVPHGQHHYIPKNELASBELAAEAPL
SRGINSRTRRQNSDRTSRNVPSPNCTNTNTNNSNTSQAQSDNIDISL
LKQYLKPLSRHVESDGLIFDPAQITSRARGVAPHGNHYHFIPEYQMSLEKRIA
RIIPRYRNHWVDSRPEPSQPTPEPSPPQAPSPIDEKLVKBAVRKVGQYV
VYPEEGISRYVPKDLBETVKNLESKQESVSHLTAKENAVAPRDEYKAY
NLTEAHKALFNKNGNSDFQALDKLERINDESTNKEKLVDLLALFAPIRHBERLGP
KNSQIYEYDEVEYIAQLADKYTSDGYIFDEHDIISDEGDAYVTPMHGSHHWIGDS
LSDKERVAAQNAIKYKGLIPLPSDADVKNAPTGDSSAAIYNRVKGBKRIPLVRLPYMV
EHTVEYKGNLIIIPKDHVHNIKFAWDDHTYKAPNGYTLDELFAIKYKVEHPDERP
HSDNGWGNASEHVLGKHSEDPKNKFKADERPVETPAEPVQVETEKVEAQLEA
EVLALAKVTSLSKANATETLAGLRLNLLTIQIMDNNSIMAEAEKLLALLKSNPSSYSK
EKIN"
complement(8049..9782)
/gene="Sp1176"
CDS
complement(8049..9782)
/gene="Sp1176"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAK75285.1"
/db_xref="GI:14972656"
/translation="MTEMILKGIASDGVAVAKAYLLVOPDLSPTETIVEDTNAEEARL
DAALQASQELSIVREKAVCTLGEAAQVDFDLHMLADPEMISQIKETIRAKKNVAB
AGLKEVDFITIFEGMEDPNYQERADIRVTKRVLANLLKGLPNPASINEEVIV
IAHDLTPSDTAQJLDKNPVRKAVFTNIIGRTSHSAIMARTUEIAAVLGTNNIIEIVKDG
ILAVNGITGEVILNPTDEQAAEFKAAGEAYAKQEMALLKDAQTVTADGKHPFLAAN
IGTPKQVGVWNNNAEAVGLYRTFELYMSQDPTDEQVEAYKAVLGEWNGKPVVVR
TMDIGDGLPELVPDMPHENNPFLGFRALRISISETGDMPRTQIRALLRASVHGQARI
MFMVALLKEFRAAKAVFDEKANILLAEGVAVADNTQVGMIELPAAMLADQFAKEV
DIFSIGTNDIQTMAADRMRQSVSLYQYFNFSILKRLNNVIAAHAEKGWAGMGE
MAGDQKAVPLLVGMGLDDEFSMSVLTIRTSMLMKLDTAKOMEYANRAUTECSTMBEV
LELQKGVNFD"
complement(9788..10051)
/gene="Sp1177"
CDS
complement(9788..10051)
/gene="Sp1177"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAK75286.1"

```

```

/db_xref="GI:14972657"
/translation="WASKDFHVVAETGIIHARPATILLVOTASKPASDITILEYKSKSVNL
KSMGVMSLGVGGADVTIISABGADDAIAIALSEWKEGLA"
Query Match 100.0%; Score 2388; DB 1; Length 10256;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
ORIGIN
QY 1 TTCCTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTCTTCTTA 60
Db 7737 TTCCTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTCTTCTTA 7678
QY 61 TATAGATGCAAAACAGACGACGCAAAAACGGAGAAATTTGACTCTCTGATGAGGTAGCAA 120
Db 7677 TATAGATGCAAAACAGACGACGCAAAAACGGAGAAATTTGACTCTCTGATGAGGTAGCAA 7618
QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 7617 GCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 7558
QY 181 TTCACATGCGACCACTATCATTAATGCTAAGTTCCCTTATGACGCTATCATCAG 240
Db 7557 TTCACATGCGACCACTATCATTAATGCTAAGTTCCCTTATGACGCTATCATCAG 7498
QY 241 TGNAGAAATTAATCATGAAGATCCAACTAATTAAGCTTAAAGATGAGGATATTTTAAATGA 300
Db 7497 TGAAGAAATTAATCATGAAGATCCAACTAATTAAGCTTAAAGATGAGGATATTTTAAATGA 7438
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGC 360
Db 7437 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGC 7378
QY 361 TGCCACCGCGGATAACGTCGGTCAAAAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG 420
Db 7377 TGCCACCGCGGATAACGTCGGTCAAAAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG 7318
QY 421 TCAACATCTGTAAGGTGGAACTCCAAAGAACGATGCTGCTGTTGCTTGCACGTTTCGCA 480
Db 7317 TCAACATCTGTAAGGTGGAACTCCAAAGAACGATGCTGCTGTTGCTTGCACGTTTCGCA 7258
QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 7257 AGGACGCTATCTACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGGATAC 7198
QY 541 TGGTGAATGCTTATATCGTTCCTCATCGAGATCATATTACCATTACCTTAAAGATGAGTT 600
Db 7197 TGGTGAATGCTTATATCGTTCCTCATCGAGATCATATTACCATTACCTTAAAGATGAGTT 7138
QY 601 ATCAGCTAGCGAGTTGCTGCTGCAGAGCCTTCTTATCTGTCGAGGAAATCTGTCAA 660
Db 7137 ATCAGCTAGCGAGTTGCTGCTGCAGAGCCTTCTTATCTGTCGAGGAAATCTGTCAA 7078
QY 661 TTCAAGAACCTATCGCGCGCAAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 7077 TTCAAGAACCTATCGCGCGCAAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 7018
QY 721 TGTAAGCAATCCAGGAACCTAATAACTAACAAGCAACAAGCAACAAGCAACAAGCAAGTCA 780
Db 7017 TGTAAGCAATCCAGGAACCTAATAACTAACAAGCAACAAGCAACAAGCAACAAGTCA 6958
QY 781 AGCAAGCTCAAGTAATGACATTCATGATGCTCTTTGAAACAGCTCTACAACTGCTTTGAG 840
Db 6957 AGCAAGCTCAAGTAATGACATTCATGATGCTCTTTGAAACAGCTCTACAACTGCTTTGAG 6898
QY 841 TCAACGACATGTGAAGATCTGATGGCCTTGTCTTTTGTATCCAGCACAAATCAAGTCAAC 900
Db 6897 TCAACGACATGTGAAGATCTGATGGCCTTGTCTTTTGTATCCAGCACAAATCAAGTCAAC 6838
QY 901 AGCTAGAGGTGTTGAGTGCACACCGAGATCATATTACGATTCATCCCTTACTCTCAAT 960
Db 6837 AGCTAGAGGTGTTGAGTGCACACCGAGATCATATTACGATTCATCCCTTACTCTCAAT 6778

```

Qy	961	GTCTGAATTTGGAAGAAAGAAATCGCTCGTATTAATTTCCCTTCGTGTATCGTTTCAAAACCAATG	1020
Db	6777	GTCTGAATTTGGAAGAAAGAAATCGCTCGTATTAATTTCCCTTCGTGTATCGTTTCAAAACCAATG	6718
Qy	1021	GGTACCAGATTTCAAGGCCAGAAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG	1080
Db	6717	GGTACCAGATTTCAAGGCCAGAAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG	6658
Qy	1081	CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT	1140
Db	6657	CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT	6598
Qy	1141	ACGAAAGTTGGGGAAGGATATGATTCGAGAAAGGGCATCTCTCGTTATGTCCTTTGC	1200
Db	6597	ACGAAAGTTGGGGAAGGATATGATTCGAGAAAGGGCATCTCTCGTTATGTCCTTTGC	6538
Qy	1201	GAAGAATTTACCATCTGAAACGTGTTAAAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG	1260
Db	6537	GAAGAATTTACCATCTGAAACGTGTTAAAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG	6478
Qy	1261	TGTTTCACACATTTAACTGCTTAAAAAGAAAATGTTGCTCTCGTGACCAAGAAATTTTA	1320
Db	6477	TGTTTCACACATTTAACTGCTTAAAAAGAAAATGTTGCTCTCGTGACCAAGAAATTTTA	6418
Qy	1321	TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCCTGTTGNAANAATAGGCTCG	1380
Db	6417	TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCCTGTTGNAANAATAGGCTCG	6358
Qy	1381	TAATTCGTATTTCCAAGCCTTAGACAATAATTTAGAACGGTTGAATGATGAATCGACTAA	1440
Db	6357	TAATTCGTATTTCCAAGCCTTAGACAATAATTTAGAACGGTTGAATGATGAATCGACTAA	6298
Qy	1441	TAAAGAAAAATTTGGTAGATGATTTATTTGGCATCTCCTAGCAACCAATACCCTCCAGAGCG	1500
Db	6297	TAAAGAAAAATTTGGTAGATGATTTATTTGGCATCTCCTAGCAACCAATACCCTCCAGAGCG	6238
Qy	1501	ACTTGGCAAAACCAAAATTTCTCAAATTTAGTATTACTGGAAGACGAAGTTCGTATTTGCTCAATT	1560
Db	6237	ACTTGGCAAAACCAAAATTTCTCAAATTTAGTATTACTGGAAGACGAAGTTCGTATTTGCTCAATT	6178
Qy	1561	AGCTGATAAGTATACACAGTCAGATGCTTACATTTTGTGACACATGATATATCACTGA	1620
Db	6177	AGCTGATAAGTATACACAGTCAGATGCTTACATTTTGTGACACATGATATATCACTGA	6118
Qy	1621	TGAAGGAGATGTCATATGTAACGCCCTCATATGGGCCATAGTCACCTGGAATTCGAAAAAGATAG	1680
Db	6117	TGAAGGAGATGTCATATGTAACGCCCTCATATGGGCCATAGTCACCTGGAATTCGAAAAAGATAG	6058
Qy	1681	CCYTTCGTGATAAGGAAAAAGTTGCAGCTCAAGCCTTACTTAAGAAAGGTTATCTTACC	1740
Db	6057	CCYTTCGTGATAAGGAAAAAGTTGCAGCTCAAGCCTTACTTAAGAAAGGTTATCTTACC	5998
Qy	1741	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTCACAGCACTATTTA	1800
Db	5997	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTCACAGCACTATTTA	5938
Qy	1801	CAATCGTGTGAAAGGGGAAAAAGCAATTCATCGTTTCGACTTCCATATATGTTGAGCA	1860
Db	5937	CAATCGTGTGAAAGGGGAAAAAGCAATTCGACTTCCATATATGTTGAGCA	5878
Qy	1861	TACAGTTGAGGTTTAAAAACGGTAAATTTGATTTTCTCATAGGATCAATTTACCATAATAT	1920
Db	5877	TACAGTTGAGGTTTAAAAACGGTAAATTTGATTTTCTCATAGGATCAATTTACCATAATAT	5818
Qy	1921	TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTTCCAAATGGGCTATACCTTGAAGA	1980
Db	5817	TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTTCCAAATGGGCTATACCTTGAAGA	5758
Qy	1981	TTTGTGTTGGCAGCAATTAAGTACTACGCTAGAACACCCCTGACGACGCTCCCAATTTCTAATGA	2040
Db	5757	TTTGTGTTGGCAGCAATTAAGTACTACGCTAGAACACCCCTGACGACGCTCCCAATTTCTAATGA	5698
Qy	2041	TGGATGGGGCAATCCCAAGTGAGCATGTGTTTAGGCAAGAAAGACCAACAGTGAAGATCCAA	2100

Db	5697	TGGATGGGCGCAATGCCAGTGAAGCATGTGTGTAGCGAAGAAAGACCAAGTGAAGATCCAAA	5638
Qy	2101	TAAGAACATTCAAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	5637	TAAGAACTTCAAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC	5578
Qy	2161	TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAAGTTTGTCTTCGAA	2220
Db	5577	TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAAGTTTGTCTTCGAA	5518
Qy	2221	AGTACGCAATCTAGCTCTGAAGCCCAAGTGCACAGAACTCTAGCTGGTTTACGAATAA	2280
Db	5517	AGTAAAGCAATCTAGCTCTGAAGCCCAAGTGCACAGAACTCTAGCTGGTTTACGAATAA	5458
Qy	2281	TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC	2340
Db	5457	TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC	5398
Qy	2341	GTTGTTAAAGGAAGTAATCTTCACTGTAGTAAGAAAAATAAAC	2389
Db	5397	GTTGTTAAAGGAAGTAATCTTCACTGTAGTAAGAAAAATAAAC	5349
RESULT 11			
LOCUS	AX571763	349980 bp	DNA linear PAT 29-NOV-2002
DEFINITION	Sequence 4982 from Patent WO02077021.		
ACCESSION	AX571763		
VERSION	AX571763.1	GI:26003955	
KEYWORDS			
SOURCE	Streptococcus pneumoniae		
ORGANISM	Streptococcus pneumoniae		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
	Streptococcus.		
REFERENCE	1		
AUTHORS	Masigiani,V., Tettelin,H. and Fraser,C.		
TITLE	Streptococcus pneumoniae proteins and nucleic acids		
JOURNAL	Patent: WO 02077021-A 4982 03-OCT-2002;		
	Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)		
FEATURES	Location/Qualifiers		
source	1..349980		
	/organism="Streptococcus pneumoniae"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:1313"		
	/note="seq 4979 too long: 2,162,598 bases-replaced by		
	following seq~seq 4979: from 0.000.001 to 0.349.980-seq		
	4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001		
	to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq		
	4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001		
	to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq		
	4986: from 2.100.001 to 2.162.598"		
ORIGIN			
Query Match	100.0%;	Score 2388;	DB 6; Length 349980;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 2388;	Conservative	0; Mismatches	1; Indels 0; Gaps 0;
Qy	1	TTCTTACGAGTTGGACTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTTA	60
Db	106950	TTCTTACGAGTTGGACTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTTA	106951
Qy	61	TATAGATGAAAAACAAGCGACGCGCAAAAAACCGAGAAATTGACTCTCTGATGAGGTTAGCAA	120
Db	106890	TATAGATGAAAAACAAGCGACGCGCAAAAAACCGAGAAATTGACTCTCTGATGAGGTTAGCAA	106831
Qy	121	CGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAGGCGTATGTCTAC	180
Db	106830	CGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAGGCGTATGTCTAC	106771
Qy	181	TTCAATGGCGACCACTATCATTAATACAAATGGTAAGGTTCTTATGACGCTATCATCAG	240
Db	106770	TTCAATGGCGACCACTATCATTAATACAAATGGTAAGGTTCTTATGACGCTATCATCAG	106711

[illegible]

GVAVPHGDHYHFIPIYSQMSLEBERIARIIPLRYSRHNHWVDSRPESQPTPEPSPSP
 POPAPNLKTDSSNSLVGYKVRKGEYVFEKIGSYRYFAKOLPSTVKNLKSJLKO
 ESVSHLTAKDSSNVAQPDQFQKAYNLLTTEAHKALFENKGRNSDFQALDKLLERLND
 ESTNKEKLVDDLLAFIAPITHPERLQKPNISOIYETEDVRIQAOLDAITTSDCGYIFDE
 HDIISDEGDAYVTPHMGHSHWIGKDSLQKAAQYATKEKGIPLSPADVKNAPT
 GUSAALYNRVKGEKKIPLVRLPYVMEHTVVKNGLIIPHDKHYHNIKFAWFDHHTY
 KAPNGVTLDELDFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDSDPNKNFKADEE
 PVETAPPEVPOVQETKEVAQKAEAVLLAKVITDSSLKANATETTLAGLRNNLTQIM
 DNNISMAEAEKILLALLGSGNPSSVSKEKIN"
 complement (8122.. .9855)
 /gene="pts1"
 /note="synonym: sprl062"
 complement (8122.. .9855)
 /gene="pts1"
 /EC number="2.7.3.9"
 /codon_start=1
 /transl_table=11
 /label=sprl062
 /product="PTS enzyme I"
 /protein_id="AAK99866.1"
 /db_xref="GI:15458684"
 /translations="MTMLKGIAGASDGVAVAKAYLVQPLSPETITVEDTNAEEARL
 DAALQADELSVIRKAVGCTLGEAAQVFDALHVLADPEMISOIKETIRAKVNAE
 AGLVEVDNFIITFEGMEDNPNYQERAAIRDTSVKVLANLLAKKKLNPASINEEIV
 IAHDLTPSDTAQDKNFVAKFVNIIGORTSHVIAARTSHVLAALVLTGNNITELVKQDV
 ILAVNGITGEVINTPDEQAEEFKAAEGAYAKQAEWALLKDAQTVTADGKHFLAAN
 IGTQVDKEVGNNGAAGVLYRYTFEFLYMSQDPDEQVEAYKAVLEGNGKPKVVR
 TMDIGDKELPYFDMPEHNPFTLGFALRISISETGDMFPTOIRALLRASVHGOLRI
 MFPVALLKEPFAAKVADNEKNALLAEGVAVADNIOVGIMIEIPAAAMLADQPAKEV
 DPFISITNDLIQITMAFDRNEQVSYQIPNPSILRLINNVIKAAHAEGKWAAGMGE
 MAGDQQAFLLVGMGLDDEFSSMSATSVLRTSLMKKLDITAKMEEYANRALTECSTMBEEV
 LEIQKEYVNFQ"
 complement (9861.. .10124)
 /gene="ptsH"
 /note="synonym: sprl063"
 complement (9861.. .10124)
 /gene="ptsH"
 /codon_start=1
 /transl_table=11
 /label=sprl063
 /product="Histidine-containing phosphocarrier protein of
 the PTS"
 /protein_id="AAK99867.1"
 /db_xref="GI:15458685"
 /translations="MASKDPHVVAETGIHARPATLLVQTASKFASDITLLEYKGSVNL
 KSMGVMSLGVGGQADVTISAEGADADDAIAAISETWEKGLA"

ORIGIN

Query Match	99.0%;	Score	2365.6;	DB 1;	Length	10320;			
Best Local Similarity	99.4%;	Pred. No.	0;						
Matches	2374;	Conservative	0;	Mismatches	15;	Indels	0;	Gaps	0;

QY	1	TTCTTAGCAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTTA	60
Db	7808	TTCTTAGCAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTTA	7749
QY	61	TATAGATGGAATAACAACGACGCAAAAACCGAGAAATTTGACTCTCATGAGGTTAGCAA	120
Db	7748	TATAGATGGAATAACAACGACGCAAAAACCGAGAAATTTGACTCTCATGAGGTTAGCAA	7689
QY	121	CGGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCA	180
Db	7688	CGGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCA	7629
QY	181	TTACATGGCGACCACTATCATTTATTAACATGGTAAGGTTCCITTTATGACGCTATCATCAG	240
Db	7628	TTACATGGCGACCACTATCATTTATTAACATGGTAAGGTTCCITTTATGATCGGATTTTCAG	7569
QY	241	TGAAGAATTACTCATGAAGAATCAAACTATAAGCTAAAGATGAGGATATGTTTAATGA	300
Db	7568	TGAAGAATTACTCATGAAGAATCAAACTATAAGCTAAAGATGAGGATATGTTTAATGA	7509
QY	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCATCTGTTTACCTTTAAGGATGC	360

Db	7508	GGTCAAGGGTGGGATATGTTTATCAAGGTAGATGAAAATACTATGTTTACCTTAAGGATGC	7449
Qy	361	TGCCCAACGGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAAACAAGACATAG	420
Db	7448	TGCCCAACGGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAAACAAGACATAG	7389
Qy	421	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGTTGGCTTGGCATCGATC	480
Db	7388	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGTTGGCTTGGCATCGATC	7329
Qy	481	AGGACGCTATCTACAGATGATGGTTATATCTTTTAAATGCTTCTGATATCATAGAGATAC	540
Db	7328	AGGACGCTATCTACAGATGATGGTTATATCTTTTAAATGCTTCTGATATCATAGAGATAC	7269
Qy	541	TGTTGATGCTTATATCGTTCTCATGGAGATCAATTACCATTAATTACCTTAAGAAATGAGTT	600
Db	7268	TGTTGATGCTTATATCGTTCTCATGGAGATCAATTACCATTAATTACCTTAAGAAATGAGTT	7209
Qy	601	ATCAGCTAGCGAGTTGGCTGTCGAGAAGCCTTCCCTATCTGGTCGAGGAAATCTGTCAAA	660
Db	7208	ATCAGCTAGCGAGTTGGCTGTCGAGAAGCCTTCCCTATCTGGTCGAGGAAATCTGTCAAA	7149
Qy	661	TTCAAGAACCTATCGCGCAAAAATAGCGATACACTTCAAGAACAAACTGGSGTACCTTC	720
Db	7148	TTCAAGAACCTATCGCGCAAAAATAGCGATACACTTCAAGAACAAACTGGSGTACCTTC	7089
Qy	721	TGTAAGCAATCCAGGAACCTACAATACTAACCAAGCAACAGCAAACTAACAGTCA	780
Db	7088	TGTAAGCAATCCAGGAACCTACAATACTAACCAAGCAACAGCAAACTAACAGTCA	7029
Qy	781	AGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAACCTGCTTTGAG	840
Db	7028	AGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAACCTGCTTTGAG	6969
Qy	841	TCAACGACATGTAGAATCTGATGGCCTTGCTTTGATCCAGCAAAATCAAGTCAAGC	900
Db	6968	TCAACGACATGTAGAATCTGATGGCCTTGCTTTGATCCAGCAAAATCAAGTCAAGC	6909
Qy	901	AGCTAGAGTGTGTCAGTGCCACACGAGATCAATTACCCTTCATCCCTTACTCTCAAA	960
Db	6908	AGCTAGAGTGTGTCAGTGCCACACGAGATCAATTACCCTTCATCCCTTACTCTCAAA	6849
Qy	961	GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCGCCCTTCGTTATCGTTCAAA	1020
Db	6848	GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCGCCCTTCGTTATCGTTCAAA	6789
Qy	1021	GGTACAGATTTCAAGGCCAGAAACAACAGTCCCAACCGACTCCCGAACTAGTCCAGG	1080
Db	6788	GGTACAGATTTCAAGGCCAGAAACAACAGTCCCAACCGACTCCCGAACTAGTCCAGG	6729
Qy	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTTGTTAGTCAAGCTGT	1140
Db	6728	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTTGTTAGTCAAGCTGT	6669
Qy	1141	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAAAGGGCATCTCGTTATGTTTTCG	1200
Db	6668	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAAAGGGCATCTCGTTATGTTTTCG	6609
Qy	1201	GAAAGATTTTACCATCTGAAACTGTTTAAAAATCTTGAAAGCAAGTATCAAAAACAAGAG	1260
Db	6608	GAAAGATTTTACCATCTGAAACTGTTTAAAAATCTTGAAAGTAAATTTATCAAAAACAAGAG	6549
Qy	1261	TGTTTTCACACATTTAACTGCTAAAAAGAAATGTTGCTCCCTGTCACCAAGATTTTA	1320
Db	6548	TGTTTTCACACATTTAACTGCTAAAAAGAAATGTTGCTCCCTGTCACCAAGATTTTA	6489
Qy	1321	TGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG	1380
Db	6488	TGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG	6429
Qy	1381	TAATTTCTGATTTTCCAAGCCCTTAGACAAATTTATAGAACCGTTGAAATGATCAATCGACTAA	1440

17550 GGGAGCGTACACAGATGATGATATATCTTTAAGCTTCTGATATCATTTGAGGACAC 17609
541 TGGTGTGCTTATATCGTTCCTCATGAGATCATTAACATTAATTCCTAAGATGAGTT 600
17610 GGGTGTGCTTATATCGTTCCTCATGAGGACCAATACCATTAATTCCTAAGATGAGTT 17669
601 ATCAGCTAGCGAGTGGCTGCTGAGAGGCTTCTATCTGGTGGAGAAATCTGTCAA 660
17670 ATCAGCTAGCGAGTGGCTGCTGAGAGGCTTCTATCTGGTGGAGAAATCTGTCAA 17729
661 TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACATTTCAAGAACAAATCGGTACCTTC 720
17730 TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACATTTCAAGAACAAATCGGTACCTTC 17789
721 TGTAAAGCAATCCAGAACTCAAAATATTAACAAGCAAGCAACCAAGCAACATCAACAGTCA 780
17790 TGTAAAGCAATCCAGAACTCAAAATATTAACAAGCAAGCAACCAAGCAACATCAACAGTCA 17849
781 AGCAAGTCAAAAGTAATGACATTTGATGCTCTTGAACAGCTCTACAACTGCCTTTGAG 840
17850 ARCAAGTCAAAAGTAATGACATTTGATGCTCTTGAACAGCTCTACAACTGCCTTTGAG 17909
841 TCAACGACATGATGATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAAAGTCCGAAC 900
17910 TCAACGACATGATGATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAAAGTCCGAAC 17969
901 AGCTAGAGGTGTGAGTGGCCACAGCGAGATCAATACCATTTCACTCTCTCAAAAT 960
17970 AGCTAGAGGTGTGAGTGGCCACAGCGAGATCAATACCATTTCACTCTCTCAAAAT 18029
961 GTCTGAATTCGAAGAAAGAAATCGCTGATTTATTTCCCTCTGTTATCGTTCAAACTATG 1020
18030 GTCTGAATTCGAAGAAAGAAATCGCTGATTTATTTCCCTCTGTTATCGTTCAAACTATG 18089
1021 GGTACCAATTCGAAGAAAGAAATCGCTGATTTATTTCCCTCTGTTATCGTTCAAACTATG 1080
18090 GGTACCAATTCGAAGAAAGAAATCGCTGATTTATTTCCCTCTGTTATCGTTCAAACTATG 18149
1081 CCCGCAACTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 1140
18150 CCCGCAACTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 18209
1141 ACGAAAGTTCGGAAGGATGATTTTCGAAGAAAGGCACTCTCGTTATGCTTTGTC 1200
18210 ACGAAAGTTCGGAAGGATGATTTTCGAAGAAAGGCACTCTCGTTATGCTTTGTC 18269
1201 GAAAGATTTACCATCTGAACTGTTTAAATCTTTGAAAGCAAGTTTATCAAAACAGAGAG 1260
18270 GAAAGATTTACCATCTGAACTGTTTAAATCTTTGAAAGCAAGTTTATCAAAACAGAGAG 18329
1261 TGTTCACACATTTAACTGCTTAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
18330 TGTTCACACATTTAACTGCTTAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 18389
1321 TGATAAGCATATATCTTTAACTGAGCTCATAAAGCCTTGTGTAATTAAGGGTCG 1380
18390 TGATAAGCATATATCTTTAACTGAGCTCATAAAGCCTTGTGTAATTAAGGGTCG 18449
1381 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAAAGCCTTGAATGATGATCACTTA 1439
18450 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAAAGCCTTGAATGATGATCACTTA 18509
1440 ATAAAGAAAATTTGGTAGATTTTATGGGATCTCTAGACCAATTCCTCCAGAGC 1499
18510 ATAAAGAAAATTTGGTAGATTTTATGGGATCTCTAGACCAATTCCTCCAGAGC 18569
1500 GACTTGGCAACCAAAATCTCAAAATTTAGTATCTGAGACCAAGTCTGATTTGCTCAAT 1559
18570 GACTTGGCAACCAAAATCTCAAAATTTAGTATCTGAGACCAAGTCTGATTTGCTCAAT 18629
1560 TAGCTGATAAGTATCAACGCTCAGATGGTTACATTTTGTATGAACATGATATCAAGT 1619

18630 TAGCTGATAAGTATCAACGCTCAGATGGTTACATTTTGTATGAACATGATATCAAGT 18689
1620 ATGAAGAGATGCATATGTAAAGCCTCATATGGGCCATAGTCACTGATTTGGAAAGATA 1679
18690 ATGAAGAGATGCATATGTAAAGCCTCATATGGGCCATAGTCACTGATTTGGAAAGATA 18749
1680 GCCTTTCTGATAAGGAAAAGTTGCGCTCAAGCTTATCTAAAGAAAAGGATTCCTAC 1739
18750 GCCTTTCTGATAAGGAAAAGTTGCGCTCAAGCTTATCTAAAGAAAAGGATTCCTAC 18809
1740 CTCATCTCCAGACGAGATTTAAAGCAATCCAACTGGAGATAGTCAGACGATATTT 1799
18810 CTCATCTCCAGACGAGATTTAAAGCAATCCAACTGGAGATAGTCAGACGATATTT 18869
1800 ACAATCGTGTCAAGGGGAAAAGCAATTCACCTCGTTCGACTTCCATATATGTTGAGC 1859
18870 ACAATCGTGTCAAGGGGAAAAGCAATTCACCTCGTTCGACTTCCATATATGTTGAGC 18929
1860 ATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTATAAGGATCATTAACATATA 1919
18930 ATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTATAAGGATCATTAACATATA 18989
1920 TTAATTTGCTTGTGTTTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAG 1979
18990 TTAATTTGCTTGTGTTTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAG 19049
1980 ATTTGTTGCGACGATTAAGTACTACGTAGAACACCCCTGAGAAAGTCCATTCATTAATG 2039
19050 ATTTGTTGCGACGATTAAGTACTACGTAGAACACCCCTGAGCAAGCTCCCATTTCTAATG 19109
2040 ATGATGGGCAATGCCAGTGGATGTTTGAAGCAAGAAAGACCAAGTGAAGATCCAA 2099
19110 ATGATGGGCAATGCCAGTGGATGTTTGAAGCAAGAAAGACCAAGTGAAGATCCAA 19169
2100 ATAGAACTTCAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCGAAGTCC 2159
19170 ATAGAACTTCAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCGAAGTCC 19229
2160 CTCAGTAGAGACTCAAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTTTGTGTCGA 2219
19230 CTCAGTAGAGACTCAAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTTTGTGTCGA 19289
2220 AAGTAAACGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAATA 2279
19290 AAGTAAACGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAATA 19349
2280 ATTTGACTCTTCAAAATTTATGATTAACATATGATCATGCGAGAGCAGAAAAATTTACTTG 2339
19350 ATTTGACTCTTCAAAATTTATGATTAACATATGATCATGCGAGAGCAGAAAAATTTACTTG 19409
2340 CGTTGTTAAAGGAGTAATCTTCTCATCTGTAAGTAAGGAAAAATTAAC 2389
19410 CGTTGTTAAAGGAGTAATCTTCTCATCTGTAAGTAAGGAAAAATTAAC 19459

RESULT 14

AF340221 2535 bp DNA linear BCT 22-MAY-2001
LOCUS Streptococcus pneumoniae PhpA (phpA) gene, complete cds.
DEFINITION Streptococcus pneumoniae
ACCESSION AF340221
VERSION AF340221.1 GI:13447093
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 2535)
AUTHORS Zhang, Y., Masi, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and Green, B.A.
TITLE Recombinant PhpA protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against intranasal pneumococcal challenge
JOURNAL Infect. Immun. 69 (6), 3827-3836 (2001)

MEDLINE PUBMED REFERENCE AUTHORS	TITLE JOURNAL	FEATURES source	gene	CDS	ORIGIN
21244685 11349048 2 (bases 1 to 2535) Zhang, Y., Masi, A., Barniak, V., Mountzouros, K., Hostetter, M. and Green, B.	Direct Submission Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA	Location/Qualifiers 1. .2535 /organism="Streptococcus pneumoniae" /mol_type="genomic DNA" /db_xref="taxon:1313" 1. .2535 /gene="phpA" 1. .2535 /gene="phpA" /note="histidine motif-containing protein" /codon_start=1 /transl_table=11 /product="phpA" /protein_id="AAK26629.1" /db_xref="GI:13447094" /translations="MKINKYLVGSAALILSCSYELGLYQARTVKENRVSVIDGK LQMKDNYLPDVEKSRGNGINAQIVIKITDQGVTSHGHDHYVNGYDAIIGSEE YATQKPNYOLKDDSEIIEIKGGVYVYVKDAHADNVRUKBINRQKSHS QHEGGTPRNDGAVALARSCRTYTDGDIYFNASDIIEDTGDAIVPHGDHYHYPKN ELSEALAAKAFLSGRNLSUNSRKTYRRQNSDNTSRNTWPSVNSGFTVINTNSNKS TNSQAQNSNDIDSLLOKYLKPLSRHVESDGLIFDPAQITTSRTANGVAVPHGDHYHF IPYSOLKPELEKLARIPLRYRNHWVPDSRPEQSPQSPTEPSPQAPNPQAPAS NPIDEKLVAERKGVGYFENGVPVRYIPAKDLGAETAAAGIDSKLAKOESLHKLG AKTDLTPSSDREYFNKAYDILLARIHQDLNKGROVDFFALDNLRLKQVSDSKVKL VDLILAFLAIRHPRLRGKNAOITYDDEILOVAKLAGKYTTEDGYIFOPRDI TSDEG DAVYTHMTHSHWIKDLSLESEBAAAQAYAKEGLTPPSTDHQDSGNTAEKGAERLY NRVKAAKVPLDRMPYNLQTVKEKNSLIIIPHYDHIHKIFWFDEGLYEAPKGYSL EDLLATYVIEVHNPSHDNGFGNASDHVQNRKNQADNTQTEKPNKEKQTEKPE EETPREKPOSEKPEPPEEPSEEPSEEPQVETEKVKEKLRLEARDLLGK IQNPITIKSNAKETLITGLIKNNLLFCTQDNNTIMAEAEKLLALLKESK"			
Query Match Best Local Similarity Matches 1724; Conservative	60.08; Score 1434.4; DB 1; Length 2535; 79.08; Pred. No. 5.6e-296; 0; Mismatches 447; Indels 12; Gaps 1;				
Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGACCGTTAAGGAAAAATAATCGTGTTCCTTA	60		
Db	60	TTCTTACGAGTTGGGACTGTATCAAGCTAGACCGTTAAGGAAAAATAATCGTGTTCCTTA	119		
Qy	61	TATAGATGGAAAAACAGCGACGCGAAAAACCGAGAAATTGACTTCCTGTAGAGTTAGCAA	120		
Db	120	TATAGATGGAAAAACAGCGACGCGAAAAACCGAGAAATTGACTTCCTGTAGAGTTAGCAA	179		
Qy	121	CGCTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGACAAGCGCTATGTCAC	180		
Db	180	CGCTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAAATNACAGACAAGCGCTATGTCAC	239		
Qy	181	TTCAATGGCGACCACTATCATTTATTAACAATGGTAAGGTTTCCTTATCAAGCTATCATCAG	240		
Db	240	TTCAATGGCGACCACTATCATTTATTAACAATGGTAAGGTTTCCTTATCAAGCTATCATCAG	299		
Qy	241	TGAAGAATTACTCATGAAGATCCAAACTATAGCTAAAGATGAGGATATTGTTAATGCA	300		
Db	300	TGAAGAATTGCTGATGAAGATCCAAACTATAGCTAAAGGACGAGGATATTATCATGTCGA	359		
Qy	301	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC	360		
Db	360	AATCAAGGGTGGTATGTCGATTAAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC	419		
Qy	361	TGCCACCGCGGATAAGTCCGTCAAAAGAGGGAATCAATCGACAAACCAAGAGCATAG	420		
Db	420	AGCTCATCGGGAATAATGTCGTCAAAAGAGGGAATCAATCGACAAACCAAGAGCATAG	479		
Qy	421	TCACATCGTGAAGGTGGAACTCCAAGAAACGATGCTGTGTTGCTTTGGCAGCTTCGCA	480		

QY	DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1560	DB	TCATCCAGAACGTTT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1549	QY	TATTGCTCAAT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1620	DB	AGTAGCCAAAGTTG	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1609	QY	TATTAATCAGTGTGA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1680	DB	TATAACCAAGTGTAT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1669	QY	TGGAAGAAGATAGC	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1740	DB	TAAAAAAGATAGTT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1729	QY	AGGTATCTACCTCA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1800	DB	AGGTTTGACCCCTC	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1789	QY	AGCAGCTATTATCA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1860	DB	AGAAGCTATCTCA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1849	QY	TATGGTTGAGCATAC	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1920	DB	CAATCTTCAATATC	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1909	QY	TTACCAATAATTAAT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1980	DB	TTACCAATAACATCA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
1969	QY	TACCTTGGAGATATT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
2040	DB	TAGTCCTTGAGGATC	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
2029	QY	ACATTCTAATCATGT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
2100	DB	GCAATTCAGATAAT	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	
2089	QY	TGAAGATCCAAATAA	Novel streptococcus antigens.	BD263585	1	GI:33073353	unidentified	unclassified.	1	Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and Charland, N.	Novel streptococcus antigens	Patent: JP 200253123-A 2 08-OCT-2002;	

Db 180 GCGTGAAGAAATCAACGCGGAAACAAATCGTTCATCAAGATTACGGATCAAGGTTATGTGAC 239
Qy 181 TTCATGGCGACCACTATCATTTATTAATTAAGTGAAGTTCCTTTATGACGCTATCATCAG 240
Db 240 CTCTCATGGAGACCAATATCATTTACTATTAATGGAAGTCCCTTTATGATGCATCATCAG 299
Qy 241 TGAAGAAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATTTGTAATGA 300
Db 300 TGAAGACTCTCATGAAGATCGAATTAATCAGTTGAGGATTCAGACATGTCAATGA 359
Qy 301 GGTCAAGGGTGGATGATGTTATCAAGGTAGATGGAATAATFACTATGTTTACCTTAAGGATGC 360
Db 360 AATCAAGGGTGGTATGTTATGTTAAAGTAAACGGTAAATFACTATGTTTACCTTTAAGGATGC 419
Qy 361 TGCCACGCGGATACGTCGTACAAAGAGGAATCAATCGACAAACAAAGAGCATAG 420
Db 420 AGCTCATGCGGATTAATGTCCGTACAAAGAGGAATCAATCGGCAAAACAAAGAACATAG 479
Qy 421 TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTCTGTTGCCCTTGGCAGCTTCGCA 480
Db 480 TCAGCATCGTGAAGAGGACTTCAGCAACGATGGTGGGTAGCCTTTGCACGTTTACA 539
Qy 481 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTTCTGATATCATAGAGATAC 540
Db 540 GGGACGCTACACACAGATGATGTTATATCTTCAATGATCTCTGATATCATCGAAGATAC 599
Qy 541 TGGTGATGCTTATATGTTCTCTCATGAGATCATTAACATTAATTCCTTAAGATGATTT 600
Db 600 GGGCGATGCTTATATGTTCTCTCATGAGATCATTAACATTAATTCCTTAAGATGATTT 659
Qy 601 ATCAGCTAGCGAGTTGGTCTGTCAGAGGCTTCTTATCTGTCGAGGAATCTGTCAAA 660
Db 660 ATCAGCTAGCGAGTTGGTCTGTCAGAGGCTTCTTATCTGTCGAGGAATCTGTCAAA 719
Qy 661 TTCAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 720 TTTAAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 779
Qy 721 TGTAAAGCAATCCAGGAATCAAACTACTACAAAGCAACAAAGCAACACTAACAGTCA 780
Db 780 TGTAAAGCAATCCAGGAATCAAACTACTACAAAGCAACAAAGCAACACTAACAGTCA 839
Qy 781 AGCAAGTCAAAATGACATTGATAGTCTCTTCAAAACAGCTCTACAAACCTGCCCTTTGAG 840
Db 840 AGCAAGTCAAAATGACATTGATAGTCTCTTCAAAACAGCTCTACAAACCTGCCCTTTGAG 899
Qy 841 TCAACGATGTAGAAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAACTCGAAC 900
Db 900 TCAACGATGTAGAAATCTGATGCCCTTATTTTCGACCCAGCGCAAAATCAAACTCGAAC 959
Qy 901 AGCTAGAGGTGTTGACGTGCCACAGGAGATCATTAACACTTATCCCTTACTCTCAAT 960
Db 960 CGCCAGAGGTGTAGCTGTCCCTCATGTGAACCATTAACCATTTATCCCTTATGAACAAAT 1019
Qy 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTAATTTCCCTTGTGTTATCGTTCAAAACCATTTG 1020
Db 1020 GTCTGAATTTGAAGAACGAATCGCTCGTATTAATTTCCCTTGTGTTATCGTTCAAAACCATTTG 1079
Qy 1021 GGTAACAGATTCAGGCGAGAACAAACAGTGCACACCGACTCCGGAACCTAGTCCAGG 1080
Db 1080 GGTAACAGATTCAGGCGAGAACAAACAGTGCACACCGACTCCGGAACCTAGTCCAGG 1139
Qy 1081 CCCGCACTCGACCAATCTTAAATAGACTCAA-----ATCTCTTTTGGT 1128
Db 1140 TCCGCAACCTTCACCAATCTCAACAGCTTCAAGCAATCCAATGATGAGAAATTTGGT 1199
Qy 1129 TAGTCAGCTGTACGAAAGTTGGGAGGATATGTTTTCGAGAAAGGSCATCTCTCG 1188
Db 1200 CAAGAAGCTGTTCAAAAGTAGGCGATGGTTATGTTCTTTGAGGAGATGGAGTTTCTCG 1259
Qy 1189 TTATGCTTTGCGAAGATTTACCATCTGAACTGTTTAAATACTTGAAGCAAGTTATC 1248
Db 1260 TTATATCCGACCAAGATCTTTTTCAGCAGAAACAGCAGGCGCATTTGATAGCAACTGGC 1319

Qy 1249 AAAACAAGAGAGTGTTCACACACTTTAACTGCTTAAAAAAGAAAATGTTGCTCTCTCGTA 1308
Db 1320 CAAGCAGGAAGTTTATCTATAAGCTAGGAGCTTAAGAAAATGACCTCCCATCTAGTGA 1379
Qy 1309 CCAAGAAATTTATGATAAAGCATATACTGTAACTGAGCTCATAAAGCTTGTGTTGN 1368
Db 1380 TCGAGAAATTTTCAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGA 1439
Qy 1369 AAATAAGGGTGTAAATTTCTGATTTCCAAAGCCTTAGACAAAATTAATAAGACGCTTTGAATGA 1428
Db 1440 TAATAAAGGTGCAACAAGTTGATTTTGAAGCTTTGGATAAACCTGTTGGAACGACTCAAGGA 1499
Qy 1429 TGAATCGACTAATAAAGAAAATTTGGTAGATGATTAATGGCATTCCTAGACCAATATAC 1488
Db 1500 TGTCTCAAGTGATAAGTCAAGTTAGTGGATGATTAATCTTGGCTTCTTAGCTCGGATTCG 1559
Qy 1489 CCATCCAGAGCGACTTCGCAAAACCAATTTCTCAAAATTTAGTATATCTGAAGACGAAGTTTCG 1548
Db 1560 TCATCCAGAGGTTTAGGAAAACCAATGCGCAATTTACCTACTGATGATGAGATTCGA 1619
Qy 1549 TATTGCTCAATTTAGCTGATTAAGTATACAACGCTCAGATGGTTTACATTTTGTGATGAACATGA 1608
Db 1620 AGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAAGACGGTTATATCTTTGATCTCTGTA 1679
Qy 1609 TATTAATCAGTGATGAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCACTGGAT 1668
Db 1680 TATAACAGTGATGAGGGGATGCTTATGTAACCTCCAATATGATCCCATAGCCACTGGAT 1739
Qy 1669 TGGAAAAGATAGCCTTTCTGATAAGGAAAAGTTGACGCTCAAGCCTTATCTAAAGAAAA 1728
Db 1740 TAAAAAAGATAGTTTGTCTGAGCTGAGAGCGGAGCCAGGCTTATGCTTAAAGAGAA 1799
Qy 1729 AGGTATCCTTACCTCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAATCGAGATAGTGC 1788
Db 1800 AGGTTTGACCCCTCTTTCGACAGACCATCAGGATTCAGGAAATACTCAGGCAAAAGGAGC 1859
Qy 1789 AGCAGCTATTTACAACTGTTGAAGGGGAAAACGAATTTCCACTGTTGACACTTCCATA 1848
Db 1860 AGAAGCTATCTACAAACCGGTGAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCCCTTA 1919
Qy 1849 TATGTTTGACATACAGTTCAGGTTTAAACGCTTAATTTGATTTATTCCTCATAGGATCA 1908
Db 1920 CAATCTTCAATATACTGTAGAGTCAAAAACGGTAGTTTAATCATACTCATTTATGACCA 1979
Qy 1909 TTACCATTAATTAATTTGCTTGTGTTTGAATGATCACACATACAAAGCTCCAAATGGCTA 1968
Db 1980 TTACCATTAACATCAAAATTTGAGTGGTTTGAAGAGGCTTTTATGAGGCACTTAAGGGGTA 2039
Qy 1969 TACCTTGGAGATTTGTTTGGAGCAGTTAAGTACTAGTAGAACACCTGACGAACTGCC 2028
Db 2040 TACTCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACTGCC 2099
Qy 2029 ACATTTCAATGATGATGGGCAATGCCAGTGCAGTGTGTTAGGCAAGAAAGAACCCAG 2088
Db 2100 GCATTCAGATAATGTTTGTGTAACGCTACGACCATGTTCAAGAGAAACAAATAATGGTCA 2159
Qy 2089 TGAAGATCCAAATAAGAACTTCAAGCGGATGAAGAG----- 2125
Db 2160 AGCTGATACCAATCAAAACGGAATAAACCAAGCGAGGAGAAAACCTTCAGACAGAAAAACCTGA 2219
Qy 2126 ----- 2125
Db 2220 GGAAGAAACCCCTCGAAGAGAGAAACCAAAAGCAGAAAACAGAGTCTCCAAAACCAAC 2279
Qy 2126 -----CCAGTAGAGGAAACACCTGCTGAGCAGAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2280 AGAGGAACCAAGAGAGAAATCACAGAGGAATCAGAAAGAACCTTCAGGTCGAGACTGAAAA 2339
Qy 2179 AGTAGAAGCCCAACTCAAAGAACGAGAAAGTTTGTCTTGGAAAAGTAAACGGATTTCTAGTCT 2238
Db 2340 GGTGTGAAGAAAAACTTGAGAGAGGCTGAAGATTTTACTTTGGAAAAAATCCAGGATCCAATTAT 2399

2239 GAAAGCCAAATGCAACAGAGAACTAGCTGGTTTACGAAATAATTTGACCTCTTCAAATAT 2298
 2400 CAAATGCAATGCAAGAGACTCTACAGGATTAATAATAATTTACTATTGGCACCA 2459
 2299 GGATTAACAAATAGTATCATGGCAGAGAGAGAAAAATTTACTTGGTGTAAAGAGAGTAA 2358
 2450 GGACAAACAATACTATTATGGCAGAGAGTGAAGAACTATTGGCTTTATTAAAGAGAGTAA 2519
 RESULT 17
 BD263590
 LOCUS
 DEFINITION Novel streptococcus antigens.
 ACCESSION BD263590
 VERSION BD263590.1 GI:33073358
 KEYWORDS JP 2002533123-A/7.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2647)
 AUTHORS Hamel,J., Brodeur,B.R., Pineau,I., Martin,D., Rioux,C. and Charland,N.
 TITLE Novel streptococcus antigens
 JOURNAL SHIRE BIOCHEM INC
 COMMENT OS S. pneumoniae
 PN JP 2002533123-A/7
 PD 08-OCT-2002
 PF 20-DEC-1999 JP 2000591190
 PR 23-DEC-1998 US 60/113800
 PI JOSEE HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI CLEMENT RIOUX,
 PI NATHALIE CHARLAND
 PC C12N15/02, A61K39/00, A61K39/09, A61K39/39, A61P11/00, A61P25/00,
 PC A61P27/16,
 PC A61P31/04, C07K14/315, C07K19/00, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10,
 PC C12P21/02, C12N15/00, C12N5/00
 CC Novel streptococcus antigens
 FH Key Location/Qualifiers
 FT source 1. .2647
 FT /organism='S. pneumoniae'.
 FEATURES
 source
 1. .2647
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 ORIGIN
 Query Match 57.5%; Score 1374.2; DB 6; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 4.1e-283;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATTCGTGTTTCCTA 60
 DB 104 TCGTTATGAACCTAGGTTTGCATCAAGCTCAAACTGTAAAGAAAAATAATTCGTGTTTCCTA 163
 QY 61 TATAGATGAAACAGACGACGCAAAACCGAGATTTGACTCCCTGATGAGGTTAGCAA 120
 DB 164 TATAGATGAAACAGACGACGCAAAACCGAGATTTGACTCCCTGATGAGGTTAGCAA 223
 QY 121 GGGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGGCTATGTCAC 180
 DB 224 GGGTGAAGGAATCAACGCCGCAAAATCGTCATCAAGATTACGGATCAAGGTTATGTGAC 283
 QY 181 TTCATGCGGACCACTATCATTTATCAATGATGAGGTTCCCTTATGAGGCTATCATCAG 240
 DB 284 CTCTCATGAGACCAATTTATCTACTATAATGCAAGGTCCTTATGATGCCATCATCAG 343
 QY 241 TGAAGAAATTTACTCATGAAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAATCA 300
 DB 344 TGAAGAGCTCCTCATGAAAGATCCGAATTTATCAAGGATTGAGGATTCAGACATTGTCAATGA 403

QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGTAGGAAAAATCTATGTTTACCTTTAAGGATGC 360
 DB 404 AATCAAGGGTGGTATGTTATGTTATTAAGGTAAACGGTAAATCTATGTTTACCTTTAAGGATGC 463
 QY 361 TGCCACCGGGATTAACGTCGCTGTAACAAAGAGGAATCAATCGACAAAACAGAGACATAG 420
 DB 464 AGCTCATCGGATAATGTCGCTACAAAAGAGGAATCAATCGGCAAAAACAGAGACATAG 523
 QY 421 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACCATGGTGTCTGTTGCTTGGCACGTTGCA 480
 DB 524 TCAGCATCGTGAAGAGGAGACTTCAGCAAAACGATGGTGGGTAGGCTTTGACGTTTACA 583
 QY 481 AGGACGCTATACACAGATGATGTTTATCTTCTGATATCATAGAGATAC 540
 DB 584 GGGACGCTACACACAGATGATGTTTATCTTCAATGATCTGATATCATCGAAGATAC 643
 QY 541 TGGTATGCTTATATCGTTCCTCATGGAGATCAATACCATTACATTCCTTAAGAATGAGTT 600
 DB 644 GGGCGATGCTTATATCGTTCCTCATGGAGATCAATACCATTACATTCCTTAAGAATGAGTT 703
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCCTTCTCTATCTGCTGAGGAAATCTGTCAAA 660
 DB 704 ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCCTTCTCTATCTGCTGAGGAAATCTGTCAAA 763
 QY 661 TTCAAGAACCTATCCCGACAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 720
 DB 764 TTTAAGAACCTATCCCGACAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 823
 QY 721 TGTAGCAATCCAGAACTACAAATTAACACACAGCAACACACACACACACACACACACAC 780
 DB 824 TGTAGCAATCCAGAACTACAAATTAACACACAGCAACACACACACACACACACACACAC 883
 QY 781 AGCAAGTCAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 DB 884 AGCAAGTCAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 943
 QY 841 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 DB 944 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1003
 QY 901 AGCTAGAGGTGTTGCAAGTCCACACAGGAGATCAATACCATTCATCCCTTACTCTCAAT 960
 DB 1004 CGCCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
 QY 961 GTCTGAATGGAAGAAACGAATCGCTGCTATTTATTCCTTCTGTTATCGTTTCAACCATG 1020
 DB 1064 GTCTGAATGGAAGAAACGAATCGCTGCTATTTATTCCTTCTGTTATCGTTTCAACCATG 1123
 QY 1021 GGTACAGATTCAAGCCAGAAACACCAAGTCCACACAGCTCCGGAACCTAGTCCAGG 1080
 DB 1124 GGTACAGATTCAAGCCAGAAACCAAGTCCACACAGCTCCGGAACCTAGTCCAGG 1183
 QY 1081 CCGCAACCTGCAACCAATCTTAAATAGACTCA-----ATTCTTCTTTGGT 1128
 DB 1184 TCCGCAACCTGCAACCAATCTTAAATAGACTCA-----ATTCTTCTTTGGT 1243
 QY 1129 TAGTCAGCTGTTACAAAAGTTGGGGAAGGATGATGATGATGATGATGATGATGATGATGAT 1188
 DB 1244 CAAAGAGCTGTTTCAAGAGTAGGCGATGTTATGTTTGGAGAGATGGAGTTTCTCG 1303
 QY 1189 TTATGTTTGGGAAAGATTTACCAATCTGAAACCTGTTTAAATCTTTAAAGCAAGTTATC 1248
 DB 1304 TTATATCCCGCAAGAAATCTTTTCAGCAAGAAACAGCAGCAGGCAATGATGAGAAACCTGC 1363
 QY 1249 AAAACAAGAGAGTGTTCACACACTTTAACTCTAAAAGAAATGTTGCTCTCGTCA 1308
 DB 1364 CAAAGCAAGAGGTTTATCTCTAAGCTAGGAGCTAAGAAACCTGACCTCCCATCTAGTGA 1423
 QY 1309 CCAAGAAATTTTATGATTAAGCATATAATCTGTTAACTAGGCTCATAAAGCCCTGTTTGN 1368
 DB 1424 TCGAGAAATTTTCAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGA 1483
 QY 1369 AAATAGGGTCTGTAATCTGATTTCCAGCCTTAGACAAATTTATTAGACGCTTGAATGA 1428

1484 TAATAAGGTCGACAAAGTTGATTTTGGGCTTTGATTAACCTGTTGGACACTCAAGGA 1543
 1429 TGAATCGACTAATTAAGAAATTTGGTAGATGATTTATTTGGATTCCTAGCACCAATTAAC 1488
 1544 TGTCTCAAGTGATTAAGTCAAGTTAGTGGATGATATTTCTTGGCTTCTTAGCTCGGATTCG 1603
 1489 CCATCCAGAGCGACTTGCACAAACCAATTTCTCAAAATGAGTATACCTGAAGACGAAGTTTCG 1548
 1604 TCATCCAGAGCTTTAGGAAACCAATTCGCAATTAACCTACCTGATGATGATTAACA 1663
 1549 TATTGCTCAATAGCTGATAGTATACAACTGATGTTTACATTTTGGATTAACATGA 1608
 1664 AGTAGCCAAGTTGCGAGCAAGTACACACAGAGAAGGTTATATCTTTGATCTCTCGTGA 1723
 1609 TATAATCAGTGATCAAGAGATGATATGTAACGCTCATATGGGCCCATAGTCACTGAT 1668
 1724 TATAACAGTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1783
 1669 TGGAAAGATAGCCCTTCTGATAAGGAAAGTTGACGCTCAAGCTCATATCTAAAGAAA 1728
 1784 TAAAAAGATAGTTTGTCTGAGCTGAGAGCGCGACGCCAGGCTTATGCTAAAGAGAA 1843
 1729 AGTATCCTACCTCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGC 1788
 1844 AGGTTGACCCCTCCTTCGACAGACCATCAGATTTCAGGAATTAATGAGGCAAAAGGAGC 1903
 1789 AGCAGCTATTACATCTGTTGAAGGGAAGAAAGAAATCCAACTCGTTGACACTTCCATA 1848
 1904 AGAAGCTATCTACAAACCGCTGAAAGCAGCTTAAGAAGGTGCGCACTTATGCTATGCTCTTA 1963
 1849 TATGGTTGAGCATACAGCTGAGGTTTAAAAACGCTAAATTTGATTATCTCTATAGGATCA 1908
 1964 CAATCTCAATATCTAGAGTCAAAACGCTGATTTTAACTCAATCACTCAATATGACCA 2023
 1909 TTACATAATATTAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTA 1968
 2024 TTACATAATCAATCAAAATTTGAGTGGTTGACGAAGGCTTTATGAGGCACCTTAAGGGGTA 2083
 1969 TACCTTGGAGATTTGTTGCGAGATTAAGTACTAGTACGAGACACCTGACGACGCTCC 2028
 2084 TACTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTGAAACATCCAAACGAACTCC 2143
 2029 ACATTTCAATGATGATGGGCAATGCGAGTACGATGTTAGGCAAGAAAGACCAACAG 2088
 2144 GCATTCAGATAATGTTTGTAACTGCTAGGACCATGTTCAAGAAACAAATAATGCTCA 2203
 2089 TGAAGTCCAAATAAGAACTTTCAAGCGGATGAAGAG----- 2125
 2204 AGCTGATACCAATCAAAACGGAAAAACCAAGCGAGAGAAACCTCAGACAGAAAAACCTGA 2263
 2126 ----- 2125
 2264 GGAAGAAACCCCTCGAGAGAGAGAAACCAAAAGCGAGAAACCGAGTCTCCAAAAACCAAC 2323
 2126 -----CCAGTAGAGAGAAACACCTGCTGAGCGAGAGAGTCCCTCAAGTAGAGACTGAAAA 2178
 2324 AGAGAACACAGAGAGAAATCACAGAGGATCAGAGAACTCAGTTCAGACTGAAAA 2383
 2179 AGTAGAAGCCCACTCAAGAGAGAGAGTTTGTCTTGGAAAAAGTAAACGGAATCTAGTCT 2238
 2384 GGTGTAAGAAAACTGAGAGAGGCTGAAGATTTACTTTGGAAAAATCCAGGATCCAAATTA 2443
 2239 GAAAGCCAAATGCAACAGAACTCTAGCTGTTTACGAAATTAATTTGACTCTTCAATTA 2298
 2444 CAGTCAATGCAAGAGAGACTCTCAGAGATTAATAAATAATTTACTATTTGGCACCCA 2503
 2299 GGATAACAATAGTATCATGCGAGAGAGAAAAATTTACTTGGCTGTTGTTAAAAAGGAAGTAA 2358
 2504 GGACAACTACTATTATGCGAGAGAGCTGAAACACTATTGGCTTTTATTAAGAGAGATTA 2563

AX343073
 LOCUS AX343073 2647 bp DNA linear PAT 12-JAN-2002
 DEFINITION Sequence 4 from Patent WO0198334.
 ACCESSION AX343073
 VERSION AX343073.1 GI:18152271
 KEYWORDS
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE
 1 Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
 Streptococcus antigens
 Patent: WO 0198334-A 4 27-DEC-2001;
 SHIRE BIOCHEM INC. (CA)
 FEATURES
 source
 1. 2647
 /location/Qualifiers
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /note="BVH-11"
 ORIGIN
 Query Match 57.5%; Score 1374.2; DB 6; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 4.1e-283;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATTAATCGTTTCTTA 60
 DB 104 TCGTTATGAACTAGTGTTCATCAAGCTCAACTGTAAAGAAAAATTAATCGTTTCTTA 163
 QY 61 TATAGATGGAAGAAACCAACGCGACGCAAAACCGAGAAATTTGACTCCTCATGAGGTTAGCAA 120
 DB 164 TATAGATGGAAGAAACCAACGCGACGCAAAACCGAGAAATTTGACTCCTCATGAGGTTAGCAA 223
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAACGAGCTATGTCA 180
 DB 224 GCGTGAAGGAATCAACGCGCAACAAATCGTCATCAAGATTACGATCAAGGTTATGTAC 283
 QY 181 TTCAATGCGGACCACTATCATTTATTAATGTTAAGGTTCCCTTATGACGCTATCATCAG 240
 DB 284 CTCTCATGAGAGCACTTATCATTAATAATGGAAGGTCCTTATGATGCCATCATCAG 343
 QY 241 TGAAGAAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA 300
 DB 344 TGAAGACTCTCATGAAGATCCGAATTAATCAGTTGAAGGATTCAGACATTTGCAATGA 403
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGGATGC 360
 DB 404 AATCAAGGGTGGTTATGTCATTAAAGGTAAACGGTAAATACTATGTTTACCTTAAAGGATGC 463
 QY 361 TGCCCAAGCGGATTAACGTCGTAACAAAGAGGAAATCAATTCGACAAAAACAAGAGCATAG 420
 DB 464 AGCTCATGCGGATTAATGTCGTAACAAAGAGGAAATCAATTCGCGCAAAAAACAAGACATAG 523
 QY 421 TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGTTGTTGCTTGGCAAGTTGCGCA 480
 DB 524 TCAGATCGTGAAGGAGGAGCTTCAGCAACGATGTTGCGGTAGCCCTTTGCAAGTTTCA 583
 QY 481 AGGACGCTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 DB 584 GGGACGCTACCAACACAGATGATGTTTATCTTCAATGATCTGATATCATCGAGATAC 643
 QY 541 TGTGATGCTTATATGTTTCTCATGAGATCATTAACCATTAATTCCTTAAGATGAGTT 600
 DB 644 GGGCGATGCTATATGTTTCTCATGAGATCATTAACCATTAATTCCTTAAGATGAGTT 703
 QY 601 ATCAGCTAGCGACTTCGCTGCTCAGAACCTTCTCTATCTGTCGAGGAAATCTGTCAA 660
 DB 704 ATCAGCTAGCGAGTTGGCTGCTCAGAACCTTCTCTATCTGTCGCGGAAATCTGTCAA 763
 QY 661 TTCAAGAACCTATCGCGCAAAAAATAGCGATAACACTTCAAGAAACAACTGGGTACTTTC 720
 DB 764 TTTAAGAACCTATCGCGCAAAAAATAGCGATAACACTTCAAGAAACAACTGGGTACTTTC 823

PC C12N5/10, C12N9/48, C12P21/08, C12N15/00, C12N5/00, A61K37/02 CC
Human complement C3-degrading polypeptides from streptococcus CC

Location/Qualifiers

FEATURES	Location/Qualifiers
source	1. .2478

ORIGIN

Query Match	42.4%;	Score 1011.8;	DB 6;	Length 2478;
Best Local Similarity	66.2%;	Pred. No. 1.1e-205;		
Matches 1628; Conservative	0;	Mismatches 688;	Indels 144;	Gaps 6;

QY	1	TTCTTAGAGTGTGGACTGTTATCAAGCTAGACCGGTTTAAGGAANA---TAATCGTGTTTC	57
Db	60	TTCTTAGAATTGGTCTGCCAACAGCTGGTTCAGGTTAAGAAGAGAGTCTAATCGAGTTTC	119
QY	58	CTATATAGATGGAAAAACAAGCGACGCAAAAATAAATTTGACTCCTCATGAGGTTAG	117
Db	120	TTATATAGATGGTGCATCAGGCTGGTCAAAGGCGAGAACTTGACACCAGATGAGTCAG	179
QY	118	CAAGCGTAGAGGAATCAATCGTGAAGAAATCGTCATCAAGATAACAGACCAAGGCTATGT	177
Db	180	TAAAGAGGAGGGGATCAACCGCGAAACAATCGTCATCAAGATTACGGATCAAGGTTATGT	239
QY	178	CAGTTCACATGGCGCACCACTATCATTTATTAATCAATGGTAAGGTTCCCTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGACCAATTCATTACTATAATGGCAAGGTCCTTATGATGCCATCAT	299
QY	238	CAGTGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCGAATTCAGTTGGAAGATTCAGACATTTGCAA	359
QY	298	TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAACTACTATGTTTACCTTAAGGA	357
Db	360	TGAAATCAAGGGTGGTTATGTTTATCAAGGTAGATGGAAAACTATGTTTACCTTAAGGA	419
QY	358	TGCTGCCACCGCGATAACGTCGTCAAAAGAGGAATAACAATCGACAAAAACAAGGCA	417
Db	420	TGCAGTCTATGCGGATAAATTCGCGACAAAGAGAGATTTAAACGTCAGAAGCAGGAACA	479
QY	418	TAGTCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGTTGCCCTTGCACGTTTC	477
Db	480	CAGTCAATATCAGGGGTGGTCT-----AAGATCAAGCATGATGTTGCACCCAGAC	533
QY	478	GCAAGACGCTATPACTACAGATGATGGTTATATCTTTAATGTTCTGATATCATAGAGGA	537
Db	534	CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGCATCTGATCATCATGAGGA	593
QY	538	TACTGGTGTGCTTATATCGTTCCCTCATGGAGATCATTACCATTACATTCCTTAAGAATGA	597
Db	594	CACGGGTGATGCTTATATCGTTCCCTCAGCGCACCATTCACATTACATTCCTTAAGAATGA	653
QY	598	GTTATCAGCTAGCGAGTTGGCTGCTCAGAAAGCCTTCCTTATCTGTGTCGAGGAATCTGTC	657
Db	654	GTTATCAGCTAGCGAGTTAGTGTCTCGAAGCCTATTGGAAATGG-----	698
QY	658	AAATTCAGAAACCTATCGCGCAGAAAAATACGGATAACACTTCAANGAACAACTGGGTACC	717
Db	699	-GAAGCAGGATCTCGTCTCTTCTTCHAGTCTTAGTTTATAATGCAATTCAGCTCAACCAA	757
QY	718	TTCTGTAAGCAATCCAGGAACCTACAAATACTAAACAAGCAACAACAGCAACATAACAG	777
Db	758	GATTGTGAGAGAACCAATCT-----GACTGTCACTCCAACCTTA	797
QY	778	TCAAGCAAGTCAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTCAAACTGCTTT	837
Db	798	TCATCAAAATCAAGGGGAAACAATTCAGAGCTTTTTACGTGAATTCGTGCTTAAACCCCTT	857

```

|||||
1938 CCATTACCAATCAATCAATTTGAGTGGTTTGACGAAGCCCTTATGAGGCACCTAAGGG 1997
Qy 1966 CTATACCTTGAAGATTTTGTGGACGATTAAGTACTAGTAGAACCTCTGACGAACG 2025
Db 1998 GTATACCTTGAAGATTTTGTGGACGATTAAGTACTAGTAGAACCTCTGACGAACG 2057
Qy 2026 TCCACATTTATGATGATGGGCAATGCCAGTGAGCATGTTAGGCAAGAACCA 2085
Db 2058 TCCGATTTACGATATGTTTGGTAAACGCTAGCGACCATGTTCAAAGAAACAAATGG 2117
Qy 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATCA----- 2121
Db 2118 TCAAGCTATACCATCAAAACGAAACCAAGGAGGAAACCTCAGACAGAAACCC 2177
Qy 2122 ----- 2121
Db 2178 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGAAAGCGAGAAACCGAGTCTTCAAACCC 2237
Qy 2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCCTCAAGTAGAGACTGAATA 2178
Db 2238 AACAGAGGAACCGAAGGAATACCAAGAGATCAGAGAACCTCAGTCGAGACTGAATA 2297
Qy 2179 AGTAGAAGCCCAACTCAAAGAGCAGAGATTTTGTTCGAAAGTAACGGATTTCTAGTCT 2238
Db 2298 GGTGAAGAAACCTGAGAGAGCTGAAGATTTACTTGGAAATATCCAGGATCCAATTA 2357
Qy 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATATTTGACTCTTCAATTA 2298
Db 2358 CAAATCCCAATGCAAGAGACTCTCAGAGATTAATAATTTACTATTTGGCACCCA 2417
Qy 2299 GATTAACATATGATCATGCGCAGAGACAGAAATTTACTTGGTGTGTTAAAGGAGTA 2358
Db 2418 GGACAAACATATATTATGGCAGAGCTGAAACCTATTTGGCTTTTAAAGGAGTA 2477

RESULT 20
BD264091
LOCUS 2481 bp DNA linear PAT 17-JUL-2003
DEFINITION Streptococcus pneumoniae proteins and nucleic acid molecules.
ACCESSION BD264091
VERSION BD264091.1 GI:33073859
KEYWORDS JP 2002531055-A/27.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 2481)
REFERENCE Gilbert,C.F.G. and Hansbro,P.M.
AUTHORS Streptococcus pneumoniae proteins and nucleic acid molecules
TITLE Patent: JP 2002531055-A 27 24-SEP-2002;
JOURNAL MICROBIAL TECHNIQUES LTD
COMMENT OS Streptococcus pneumoniae
PN JP 2002531055-A/27
PD 24-SEP-2002
PF 27-JUL-1999 JP 2000562519
PR 27-JUL-1998 GB 9816337.1.19-MAR-1999 US 60/125164 PI
CHRISTOPHE FRANCOIS GUY GILBERT,PHILIP MICHAEL HANSBRO PC
C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61K48/00,A61P31/04, PC
C07K14/315,
PC C07K16/12,C12Q1/14,C12Q1/68,G01N33/53,G01N33/566,G01N33/569//
PC C12P21/08,
PC G01N33/15,G01N33/50,(C12N15/09,C12R1:46),(C12Q1/14,C12R1:46),
PC C12N15/00,
PC C12N15/00,(C12N15/00,C12R1:46)
CC Streptococcus pneumoniae proteins and nucleic acid molecules
FH Key Location/Qualifiers
FT source 1. .2481
FT /organism='Streptococcus pneumoniae'.
FT Location/Qualifiers
1. .2481
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
FEATURES
source

```

```

/db_xref="taxon:1313"
ORIGIN
Query Match 42.0%; Score 1003.8; DB 6; Length 2481;
Best Local Similarity 66.0%; Pred. No. 5.8e-204; Indels 144; Gaps 6;
Matches 1623; Conservative 0; Mismatches 693;
Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC 57
Db 60 TTCCTATGAGCTTGGTGGTCAACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119
Qy 58 CTATATAGATGGAAAAAAGCGACGCAAAAAACGGAGAAATTGACTCTCTGATGAGTTAG 117
Db 120 TTATATAGATGGTATCAGGCTGGTCAAAAGGCGAAAAAAGTTCACCCAGATGAAGTCAG 179
Qy 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATTAACAGATTAACAGACCAAGGCTATGT 177
Db 180 TAAGAGGAGGGGATCAACGCGCAACAAATCGTCATCAAGATTAACAGATTAACAGGTTATGT 239
Qy 178 CACTTCAATGCGGACCACTATCAATTAATCAATGGTAAAGTTCCTTATGACGCTATCAT 237
Db 240 GACCTCTCATGAGACCAATTAATCAATTAATGGCAAGGTCCCTTATGATGCCATCAT 299
Qy 238 CAGTGAAGAAATTAATCATGAAAGATCCAAACTAATTAAGCTAAAGATGAGGATATTGTTAA 297
Db 300 CAGTGAAGAGCTCCCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTTGTC 359
Qy 298 TGAGGTCAAGGCTGATATGTTATCAAGCTAGATGGAATAATTAATTAATGTTTACCTTAAAG 357
Db 360 TGAATCAAGGCTGTTATGATCATCAAGCTAGACGGAATAATTAATTAATGTTTACCTTAAAG 419
Qy 358 TGCTGCCACCGGATAACGTCCTGTAACAAAGAGGAAATCAATCAACAAAAACAAGACA 417
Db 420 TGCACCTCATGCGGATAATTAATTCGCAACAAAGAGAGATTAACCGTCAGAAGCAGAAAC 479
Qy 418 TAGTCAACATCGTGAAGGTGGAACCTCCAGAAACCATGGTGTGCTTGGCAGCTTC 477
Db 480 CAGTCAATTAATCAAGGCTCAGGAGCT-----AACGATCATGAGTAGTGCAGCCAGAC 533
Qy 478 GCAAGAGCGCTATATACAGATGATGGTTATATCTTTAATGCTTCTGATATATAGAGA 537
Db 534 CCAAGAGCGCTATATCAACGATGATGGGTATATCTTCAATGATCTGATATCATTTAGAGA 593
Qy 538 TACTGGTATGCTTATATGTTCTTCATGAGATCATTAACCATTAATTCCTTAAGAAATGA 597
Db 594 CACGGGTGATGCTTATATGTTCTTCACGGCGACCATTAACCATTAATTCCTTAAGAAATGA 653
Qy 599 GTTATCAGCTAGCGAGTTGGCTGCTGACAGAGCCTTCTTATCTGCTCGAGGAAATCTGTC 657
Db 654 GTTATCAGCTAGCGAGTTAGCTGCTGACAGAGCCTTATTTGGAATGG----- 698
Qy 658 AAATTCAGAAACCTATCGCGGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db 699 -GAAGCAGGATCTCGTCTTCTTCAAGTTCTAGTTATTAATGCAATCCAGCTCAACCAA 757
Qy 718 TTCTGTGAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAG 777
Db 758 GATTTGTCAAGAAACCAACAATCT-----GACTGTCACTCCAATTA 797
Qy 778 TCAAGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAACTGCCTTT 837
Db 798 TCATCAAAATCAAGGGGAAAAACATTTCAAGCCCTTTTACGTGAATGTTGTTAAACCCCTT 857
Qy 838 GAGTCAACACACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCACAAAATCAACAGTCG 897
Db 858 ATCAGAACCCATGTGGATCTGATGGCTTTATTTTCGACCCAGCGCAAAATCAACAGTCG 917
Qy 898 AACAGCTAGAGTGTGAGTGCCACACGAGATCATTAACCATTTACCTTCTTCTTCTCTCA 957
Db 918 AACCCGAGAGGTGATGCTGCTCCCTCATGTTAAACCATTTATCCCTTTATGAAACA 977
Qy 958 AATGCTGAATTTGAAGAACGAATCGCTGATATATTCCTTCTGTTTATCGTTTCAACCA 1017

```

Db 978 AATGCTCTGAATTTGGAAGAAACGAATTTGCTGATTAATTTATTTCCCTTCTGTTATCGTTCAAAACCA 1037
Qy 1018 TTGGGTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCACCAACCGACTCGGACCACTAGTCC 1077
Db 1038 TTGGGTACCAAGATTCAAGACAGAACCAACCAAGTCCACCAACCGACTCGGACCACTAGTCC 1097
Qy 1078 AGGCCCGCAACCTGACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTT 1125
Db 1098 AAGTCCGCAACCTGACCAAAATCTCAACAGCTCCAAAGCAATCCAATTGATGAGAAATT 1157
Qy 1126 GGTATTAGTCTGTTAGCAAAAGTTGGGGAAGATATGATTTCAAGAAAGGCGCATCTC 1185
Db 1158 GGTCAAAAGAGCTGTTTCAAAAGTAGCGATGTTTATGTTCTTTGAGGAGATGAGTTTC 1217
Qy 1186 TCGTTATGCTTTTGGCAAGATTTACCATCTGAAACTGTTAAATCTTAAATCTTGAAGCAAGTT 1245
Db 1218 TCGTTATATCCAGCCGAGGATCTTTTCAAGAAACAGCAGCGCATTTGATAGCAAACT 1277
Qy 1246 ATCAAAACAAGAGAGTGTTCACACACTTTAACTGTCTAAAGAAAGAAATGTTGCTCTCG 1305
Db 1278 GGCCAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAACTGACCTCCCATCTAG 1337
Qy 1306 TGACCAAGAAATTTATGATAAGCATATATCTGTTAACTAGGCTCATAAAGCCTTGT 1365
Db 1338 TGATCGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
Qy 1366 TGAATAAAGGGTCTGAATTTCTGATTTCCAGGCTTAGCAAAATTTAGAACGTTGAA 1425
Db 1398 TGATAATAAGGTCGACAACTGATTTGAGGCTTTGGATAACTGTTGAAACGACTCAA 1457
Qy 1426 TGATGAATCGCAATTAAGAAAGAAATTTGTTAGATGATTTTATGCGATTCCTAGCACCAAT 1485
Db 1458 GGATGTCCCAAGTGATAAGTCAAGTTAGTGGATGATATTTCTTGCTTCTTAGCTCGAT 1517
Qy 1486 TACCCATCCAGCGGACTTGGGAAACCAATTTCTCAATTTAGTATGATTAAGCAAGCAAGT 1545
Db 1518 TCGTATCCAGAAAGCTTTAGGAAACCAATTTGGCAAAATTTACCTACACTGATGATGAT 1577
Qy 1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACTGATGATGATGATGATGATGATGATGAT 1605
Db 1578 TCAAGTAGCAAGTTGGCAGGAAAGTACAAAGAGAGCGTTATATCTTTGATCCTCG 1637
Qy 1606 TGATATAATCAGTGATCAAGGAGATGATATGATTAAGCGCTCATATGGCCATAGTCACTG 1665
Db 1638 TGATATAACCAAGTATGAGGAGATGCTATGATTAACCTCAATATGACCATAGCACTG 1697
Qy 1666 GATTGAAAGATAGCTTCTGATAGGAAAGTTGAGCTCAAGCTCAAGCTTACTAAAGA 1725
Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCGCCAGGCTTATGCTAAAGA 1757
Qy 1726 AAAAGGTATCTTACCTCCATCTCCAGAGCAGATGTTAAAGCAAAATCCAACTGGAGATAG 1785
Db 1758 GAAAGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATCTAGGACCAAAAG 1817
Qy 1786 TGACAGCACTATTACAATCGTTGTAAGGGGAAAAACGAATTTCCACTCGTTTCCGCTTCC 1845
Db 1818 AGCAGAAGCTATCTACAACCGGTGAAGCAGCTAAGAGAGTCCCACTTATGATGATGCC 1877
Qy 1846 ATATATGGTTGAGCATACAGTTGAGTTTAAACCGGTAAATTTGATTTATCTTCCATAGGA 1905
Db 1878 TTACATCTTCAATATCTGTAAGTCAAAACCGGTAGTTTAACTACCTCATTTATGA 1937
Qy 1906 TCATTACCAATATTTAAATTTCTGTTGTTGATGATCACATACAAAGCTCCAAATGG 1965
Db 1938 CCATTACCATACATCAAAATTTGAGTTGGTTGACGAGGCTTTATGAGGCACTTAAGG 1997
Qy 1966 CTATACCTTGGAGATTTGTTTGGCAGGATTAAGTACTAGTAAACACCTGACCAAGC 2025
Db 1998 GTATACTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGCGAATCTCAACCAAGC 2057
Qy 2026 TCCACATTTCAATGATGGGATGGGCAATGCCAGTGAAGATGTTAGGCAAGAAAGACCA 2085
Db 2058 TCGCATTCAGATAATGTTGTTTGGTAACGTAGCGACCATGTTTCAAGAAACCAAAATGG 2117

Qy 2086 CAGTGAAGATCCAAATTAAGAACTTCAAAAGCGGATGA----- 2121
Db 2118 TCAAGCTGATACCAATCAAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
Qy 2122 ----- 2121
Db 2178 TGAGGAAGAAACCCCTCGAAGAGAAACCGAAAGCGAGAAACGAGTCTTCCAAAACC 2237
Qy 2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2238 AACAGAGAAACCAAGAAATCACAGAGGAATCAGAGAACTCAGGTCAGACTGAAAA 2297
Qy 2179 AGTAGAAGCCCACTCAAGAGAGAGAGTTTTCGTTTCGAAAGTAAGGATTTAGTCT 2238
Db 2298 GGTGGAAGAAAACTTGAGAGAGGCTGAAGATTTACTTTGGAATAATCCAGGATCCAAATTAT 2357
Qy 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATTAATTTGACTCTTCAATAT 2298
Db 2358 CAAGTCCATGCAAGAGACTCTCAGGATTTAAAAATAATTTACTATTGGCACCCA 2417
Qy 2299 GGATAACAATAGTATCATGCGAGAGCAGAAAAATTTACTTGGCTTTGTTAAAAAGGAAGTAA 2358
Db 2418 GGACAACTACTATTATGCGCAGAGCTGAABAACTATTGGCTTTATTAAGGAGAGTAA 2477

RESULT 21
AX569137
LOCUS AX569137 2457 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 2345 from Patent WO202077021.
ACCESSION AX569137
VERSION AX569137.1 GI:26002635
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1
AUTHORS Masignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2345 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES
source
1..2457
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"

ORIGIN

Query Match 41.6%; Score 993; DB 6; Length 2457;
Best Local Similarity 65.8%; Pred. No. 1.2e-201;
Matches 1611; Conservative 0; Mismatches 696; Indels 141; Gaps 6;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTCTTTC 57
Db 60 TTCTTATGAGCTTGGAGCTTACCAAGCTGCTCAGGATAAGAGAGTCTAACTCGAGTTGC 119
Qy 58 CTATATAGATGAAAAACAAGCGACCGCAAAACCGGAGAAATTTCACTCTCTGATGAGTTAG 117
Db 120 TTATATAGATGTTGATCATCGGCTGGTCAAAAGCGCAGAAACTTGACACAGATGAAGTCAG 179
Qy 118 CAAGCGTGAAGAAATCAATGCTGAGCAAAATCTCATCAAGATAACAGACCAAGGCTATGT 177
Db 180 TAAGAGGGAGGGGATCAACGCCGAAACAAATTTTATCAAGATTACGGATCAAGGTTATGT 239
Qy 178 CACTTCAATGCGGACCACTATCATTTTACAACTGTAAGGTTTCTTATGAGCTATCAT 237
Db 240 GACCTCTCATGAGACCATTAATCTATTAATGGAAGGTTCTTATGATGCCATCAT 299
Qy 238 CAGTGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATTTGTTAA 297
Db 300 CAGTGAAGAGCTCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTGACAGATTGTCA 359

QY 298 TGAGGCTCAAGGCTGGATATGTTATCAAGGTAGATGGAAATACATATGTTTACCTTAAAGGA 357
Db 360 TGAATCAAGGCTGGTATGTTATTAAGGTAAACGGTAAATACATATGTTTACCTTAAAGGA 419
QY 358 TGCTGCCACCGGATACCGTCCGTACAAAGAGAGAAATCAATCGACAAAACAAAGAGCA 417
Db 420 TGCAGCTCATCGGATATATTCGGACAAAGAGAGATTAACCGTCAGAGCAGGAACG 479
QY 418 TAGTCAACATCGTGAAGGTGAATCTCCAAAGAAACGATGGTCTGTTCCTGGCAGCGTTC 477
Db 480 CAGTCATAATCAT-----AACTCAAGAGCAGATAATGCTGTGTCGAGCCAGAGC 530
QY 478 GCRAAGGCGCTTACTACAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGGA 537
Db 531 CCAAGGCGTTATACACGGATGATGGGTATATCTTCAATGCACTGATATCATTTGAGA 590
QY 538 TACTGGTGATGCTTATATCGTTCTCATGGAGATCATPACCATTAATTCATTCCTAAGAATGA 597
Db 591 CACGGGTGATGCTTATATCGTTCTCACGGGACCATTAACCATTAACATTCCTAAGAATGA 650
QY 598 GTTATCAGTCAGGAGTTGGCTGCTGAGAGGCTTCTATCTGTCGAGGAATCTGTC 657
Db 651 GTTATCAGTCAGGAGTTAGCTGCTGCAGAGGCTTATGGAATGG----- 695
QY 658 AAATTCRAAGAACCTATCCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db 696 -GAAGCAGGATCTCGTCTTCCAGTTCTAAGTTATTAAGTCAATCCAGCTCAACCAA 754
QY 718 TTCTGTAAAGCAATCCAGGAACCTACAAATACCTAACACAAGCAACACAGCAACACTAACAG 777
Db 755 GATTGTCAGAGAACCAACATCT-----GACTGTCACCTCCAACCTTA 794
QY 778 TCRAAGCAAGTCAAGTAATGACATGATAGTCTCTTTGAAACAGCTCTACAACTGCCCTT 837
Db 795 TCATCAAAATCAAGGGGAAAACATTTCAAGGCTTTTACGTAATTTGATGTAAACCCCTT 854
QY 838 GAGTCAACGACATGATAGATCTGATGGCTCTCTTTGATCAGACACAAATCACAAAGTCG 897
Db 855 ATCAGAACGCTATGTTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAAGTCG 914
QY 898 AACAGCTAGAGGTGTTCAGTGCACACGGAGATCATTAACAATTCATCCTTACTCTCA 957
Db 915 AACGCCAGAGGTGTAGCTGCTCCATGTTAACTTACCACTTTATCCCTTATTAAGACA 974
QY 958 ATGCTGAATTTGGAAGACGATCGCTGATATATCCCTTCGTTATCGTTTCAACCA 1017
Db 975 AATGCTGAATTTGGAACGAAATGCTGATATTTCCCTTCGTTATCGTTTCAACCA 1034
QY 1018 TTGGGTACAGATTCGAAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 1077
Db 1035 TTGGGTACAGATTCGAAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 1094
QY 1078 AGGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCT 1137
Db 1095 AAGTCGGCAACCACTCCAAAGCAATCCAATTTGATGAGAA-----ATTGTCAAAAGAGC 1148
QY 1138 GGTACGAAGGTTGGGAGGATATGATTTCCGAAGAAAGGCACTCTCGTTATGTCCT 1197
Db 1149 TGTTGAAAGTAGGCGATGGTTATGCTTTGAGGAGATGAGTTTCTCGTTATATCCC 1208
QY 1198 TCGCAAGAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGA 1257
Db 1209 AGCCCAAGGATCTTCAGCAGAAACACAGCAGCGCATTTGATAGCAAACTGGCCAGCAGA 1268
QY 1258 GAGTGTTCACACATTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAATT 1317
Db 1269 AAGTTTATCTATAAGCTAGGAACCTAAGAAACCTGACCTCCCATCTAGTGATCGGAATT 1328
QY 1318 TTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTCTTTGNAATAAAGG 1377
Db 1329 TTACATAAGGCTTATGCTTACTAGCAAGAAATTCACCAAGATTTACTTGAATAAAGG 1388
QY 1378 TCGTAATCTGATTTCCAGGCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGAC 1437

Db 1389 TCGCAAGGTTGATTTTGAGGCTTTGGATAAACCTGTTGGAACGACTCAAGGATGTTCTCAAG 1448
QY 1438 TAATAAGAAATAATGGTAGATGATTTTATTTGGCAATCTCTAGCACCAATTTACCCATCCAGA 1497
Db 1449 TGATAAAGTCAAGTTAGTGGAAAGATAATCTTTCGCTTCTTAGCTCCGATTTGCTCATCCAGA 1508
QY 1498 GCAGCTTTGGCAAAACCAAAATCTCAAAATTGAGTATACCTGAAGCAAGTTTCGATTTGCTCA 1557
Db 1509 ACGTTTAGGAAACCAAAATGCGCAAAATTAACCTACATGATGAGATTCAGTAGCCAA 1568
QY 1558 ATTAGCTGATAAGTATACAAAGTCAGATGGTTAATTTTTGATGAAACATGATATAATCAG 1617
Db 1569 GTTGGCAGGCAAGTACACAACAGAAAGACGGTTATATCTTTGATCTCTGATATAAACCAAG 1628
QY 1618 TGATGAAGGAGATGCATATGTAACGCTCATATGCGCCATAGTCACTGGAATTTGGAAGAAGA 1677
Db 1629 TGATGAGGGGATGCCATATGTAACCTCAACATGACCCATAGCCACTGGATTAAGAAAGA 1688
QY 1678 TAGCCTTTCTGATAAGGAAAGTTTGCAGCTCAAGCCTTATCTAAAGAAAGATATCCT 1737
Db 1689 TAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGCTTATGCTAAAGAGAAAGGTTTGAC 1748
QY 1738 ACCTCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCGACGAGCTAT 1797
Db 1749 CCCTCCTTCGACAGACCATCAGGATTCAGGAAATACCTGAGGCAAAAGAGCAGAGCTAT 1808
QY 1798 TTACAATCGTGTGAAGGGGAAACCAAAATCCCACTCGTTCGACTTCCTCATATATGTTGA 1857
Db 1809 CTACAACCGCTGAAAGCAGCTTAAGAGTGCCACTTGTATGATGCTTACATCTTCA 1868
QY 1858 GCATACAGCTTGAGGTTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTAACATAA 1917
Db 1869 ATATCTGTGAAGTCAAAACCGTAGTTTAAATCATACCTCATATGATGACCATTAACATAA 1928
QY 1918 TATTAATTTGCTGTTTGTATGATACACATACAAAGCTCCAATGSGTATACCTTGA 1977
Db 1929 CATCAAAATTTGAGTGTGTTGACGAAGGCTTTTATGAGGCACTTAAGGGGTATACTCTTGA 1988
QY 1978 AGATTTGTTTCGACGATTAAAGTACTAGTAGAACACCTGACGAACCTGCACATTTCTAA 2037
Db 1989 GGATCTTTTGGGAGCTGTCAGAGTACTATGTCGAACATCCAAACGAAACGTCGCAATCAGA 2048
QY 2038 TGATGATGGGCAATGCGAGTGAAGTGTGTTAGGCAAGAAAGACACAGTGAAGATCC 2097
Db 2049 TAAATGTTTGTGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCAAGCTGATAC 2108
QY 2098 AAATAAGAACTTCAAAGCGGATGA----- 2121
Db 2109 CAATCAAAACGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAAC 2168
QY 2122 -----AGAGCCAGT 2130
Db 2169 CCCTCGAAGAGAAACCGAAACCGGAGAAACCAAGCTCTCCAAACCAACAGAGAAACC 2228
QY 2131 AGAGAAAAACCTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCA 2190
Db 2229 AGAAGAAATCAACAGAGGAATCAGAAAGAACCTCAGGTGAGACTGAAAAAGGTTGAAGAAA 2288
QY 2191 ACTCAAGAGCAGAAAGTTTTCCTTCGGAAGTAACGGATTTCTAGCTCTGAAGCCAAATGC 2250
Db 2289 ACTGAGAGGCTGAAGATTTTACTTTGGAAAAATCCAGGATCCAATTTATCAAGTCCAATGC 2348
QY 2251 AACGAAACTCTAGCTGGTTTACGAAATAAATTTGACTCTTCAAAATTTATGGATAACAATAG 2310
Db 2349 CAAAGAGACTCTCAGAGGATTAAGAAATTAATTTACTTATTTGCGCCACCCAGCAACAATAC 2408
QY 2311 TATCATGCGAAGCAGAAAAATTTACTTCGCGTTGTTTAAAGGAAGTAA 2358
Db 2409 TATTATGCGAAGCTGAAAAACTATTGCGCTTTTATTTAAAGGAGAGTAA 2456

RESULT 22


```

AF3118954
LOCUS       2457 bp    DNA    linear    BCT 11-FEB-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein B
            precursor (phtB) gene, partial cds.
ACCESSION  AF3118954
VERSION     AF3118954.1 GI:12744741
KEYWORDS   Streptococcus pneumoniae
SOURCE     Streptococcus pneumoniae
ORGANISM   Streptococcus pneumoniae
REFERENCE  1 (bases 1 to 2457)
            Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M. and
            Johnson, S.
            Direct Submission
            Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
            West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES   Location/Qualifiers
            1..2457
             /organism="Streptococcus pneumoniae"
             /mol_type="genomic DNA"
             /serotype="4"
             /db_xref="taxon:1313"
            1..2457
             /gene="phtB"
            1..2457
             /gene="phtB"
             /note="PhtB; protective pneumococcal antigen"
             /codon_start=1
             /transl_table=11
             /product="pneumococcal histidine triad protein B
             precursor"
             /protein_id="AAK06759.1"
             /db_xref="GI:12744742"
             /translation="MKINKKVLGASVAVLALSCVSELYGVQAGQDKKESNRVAVTDG
             DOAGOKAENLTDPDEYSKREGINAEQIVKITDQGVYVTSQDHYHYNGKVPDAIISE
             ELLMDPNVOLKDSQIDVNEIKGVYIKVNGKYYVYLKRAHADNIRKEEIKRQKQER
             SHNSRADNAVAAQAQRYTIDGIFNASDIIEDTDGAYIVPHGDHYHYPKNEI
             SASLAAEAYWNGQSGSPSSSYNAPQPRLSNHNLTVTPTHQNGENISSI
             LAELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPEQMSLEKRIA
             RIIPRYRSNHWPPDRPEEPSPQPTPEPSPQAPSPNIPDGKLKVEAVRVKGVGY
             FEENGVSRYIPAKDLISAETAAGIDSKLAKQESLSHKLGTQKTDLPSSDREFYFNKAYDL
             LARIHQDLLDNKGRQVDFEALDNLLERLKDVSDDKVLVEDILAFIAPIRHFERLGKP
             NQAIYTDDEIQVAKLAGITADGTFIDPDROITDDEGDYVTPHMTSHWIKKDSLS
             EAERAQAQYAEKGLTPSTDHQSGNTEAKGAEAIVNRVAKKVPDLRMPYNLQY
             TVEVNGSLIIPHYDHNIKFEWFDEGLPEAPKGVLTDLATLVKYVVEHPNERPHS
             DNGFGNADHVQNKQADNTQDEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEKPEK
             BEPEPESPREPQVETEKVEKLEAEADLLCKIQDPIKSNKAKETLGLKNNLLFGT
             QDNNTIMAEAKLLALLKESK"
            1..87
             /gene="phtB"
            247..264
             /gene="phtB"
             /note="Region: histidine triad"
            616..633
             /gene="phtB"
             /note="Region: histidine triad"
            718..792
             /gene="phtB"
             /note="Region: histidine triad"
            940..957
             /gene="phtB"
             /note="Region: coiled-coil domain"
            1396..1488
             /gene="phtB"
             /note="Region: histidine triad"

```

```

            /gene="phtB"
            /note="Region: coiled-coil domain"
            1657..1674
             /gene="phtB"
            1909..1926
             /note="Region: histidine triad"
            /gene="phtB"
            /note="Region: histidine triad"
            525..714
             /rpt_type=direct
            1578..1758
             /rpt_type=direct
ORIGIN
Query Match      41.5%; Score 991; DB 1; Length 2457;
Best Local Similarity 65.7%; Pred. No. 3.2e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;
QY      1  TTCTTACGAGTTGGGACTGTATCAAGCTAGACGCGTTAAGGAAAA---TAATCGTGTTC 57
DB      60  TTCTTATGAGCTTGGACGTTACCAAGCTGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119
QY      58  CTATATAGATGGAACAAACAGCGACGCAAAAACGAGAGATTTTCACCTCCTCATGAGGTTAG 117
DB     120  TTATATAGATGGTGATCAGGCTGTCAAAAGGCGAARAACTTGACACGATGAAGTCAG 179
QY     118  CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
DB     180  TAAGAGGAGGGGATCAACGCGCAACAAATGTTTATCAAGATTACGGATCAAGGTTATGT 239
QY     178  CACTTCACATGGCGACCTATCATTTATACATGTTAGGTTCCCTATGACGCTATCAT 237
DB     240  GACCTCTCATGGAGACCATTAATTAATTAATGCAAGGTTCTTATGATGCCATCAT 299
QY     238  CAGTGAAGAAATTAATCATGAAGATCCAAACTATAGCTAAAGATGAGGATATTCGTTAA 297
DB     300  CAGTGAAGAGCTCCTCATGAAGATCCGAAATTAATCAGTTGAAGGATTCAGACATTCGCA 359
QY     298  TGAGGTCAGGCTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGA 357
DB     360  TGAATCAAGGCTGGTTATGTATTAAGGTAAACGGTAATAACTATGTTTACCTTAAGGA 419
QY     358  TGCTGCCACGCGGATAACGTCGCTACAAAAGAGGAATCAATCGACAAAACAGAGCA 417
DB     420  TCGRGCTCATGCGGATAATATTCGGACAAAAGAGAGATTAACCGTCAGAACGAGAACG 479
QY     418  TAGTCAACATCTGAGGTTGAATCCAAAGAAACGATGCTGTTGCTTGCCTTGCACGTTTC 477
DB     480  CAGTCATATCAT-----AACTCAAGACAGATATGCTGTTGCTGCAGCCAGAGC 530
QY     478  GCAAGACGCTATATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
DB     531  CCAAGGACGTTATACAACGAGATGAGGGTATATCTTCAATGCATCTGATATCATTAGGA 590
QY     538  TACTGCTGATGCTTATATCGTTCCTCATCGGATCATTAACCATTAATCTTCCCTAAGATGA 597
DB     591  CACGGGTGATGCTTATATGTTCTTCACGGGACCATTAACCATTAATCTTCCCTAAGATGA 650
QY     598  GTTATCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCCTATCTGCTCGAGGAAATCTGTC 657
DB     651  GTTATCAGCTAGCGAGTTAGCTGCTGCAGAGCCTATTGGATGG----- 695
QY     658  AAATTCAAGACCTTATCGCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
DB     696  -GAAGCAGGGATCTCGTCTCTTCAAGTCTTAGTTATAATGCAATCCAGCTCAACCAA 754
QY     718  TTCTGTAGCAATCCAGGNACTACAAATACTAAACAAGACACACAGCAACACTAAGCAG 777
DB     755  GATTGTCAAGAAACCAATCT-----GACTGTCACTCCAACCTTA 794
QY     778  TCAAGCAAGTCAAGTAATGATGATGATGATGCTCTTTGAAACAGCTCTCAAACTCCCTT 837
DB     795  TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTTATGCTAAACCTTT 854

```

QY	838	GAGTCACGACATGTAGAAATCTGATGGCCTTGCTTTGATCGAGACAATCACAAAGTCG	897
Db	855	ATCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAAGTCG	914
QY	898	AACAGCTAGAGGTGTTCAGTGGCCACACGAGATCAATACCACTTCATCCCTTACTCTCA	957
Db	915	AACCGCCAGAGGTGTAGCTGTCCCTCATGTAAACCATTACCACTTTATCCCTTATGAACA	974
QY	958	AATGTCCTGAATTTGGAAGAAAGAAATCGCTCGTATATATCCCTTCGTTATCGTTCAAAACA	1017
Db	975	AATGTCCTGAATTTGAAAGAAAGAAATGCTGCTATATATCCCTTCGTTATCGTTCAAAACA	1034
QY	1018	TTGGGTACCAGATTTCAAGGCCAGAAACAAACAAGTCCACAACCGACTCCGGAACCTAGTCC	1077
Db	1035	TTGGGTACCAGATTTCAAGACACGAAGAACCAAGTCCACAACCGACTCCGGAACCTAGTCC	1094
QY	1078	AGGCCCGAACCTGACCAAAATCTTAAATPAGACTCAAAATCTTCTTTGGTTAGTCAGCT	1137
Db	1095	AAGTCCGCAACAGCTCCAAGCAATCAATTTGATGGAA-----ATTGGTCAAGAAGC	1148
QY	1138	GGTACGAAAGTTGGGGAAGGATGATTTTCGAAGAAAGGCGCATCTCTGTTATGCTCTT	1197
Db	1149	TGTTCGAAAGATGAGCGATGTTAATGCTTTTGGAGAGATGAGATTTCTCGTTATATCCC	1208
QY	1198	TGCGAAAGATTTAGCATCTGAAACTGTAAAACTGTAAAAAGCAAGTTATCAAAACAAGA	1257
Db	1209	AGCCAAGGATCTTTTCAGCAGAAAACAGCAGCAGGCATTTGATGCAAACTGCCCAAGCAGGA	1268
QY	1258	GAGTGTTCACACACTTTAACTGCTAAAAAAGAAAGATTTGCTTCCTCGTACCAACAAGATT	1317
Db	1269	AAGTTTATCTCATAGCTTAGGAATCAAGAAAACCTGACCTCCCATCTAGTATCGAGAATT	1328
QY	1318	TTATGATAAGCATATAATCTGTAACTGAGGCTCATAAAGCCCTGTGTGNAATTAAGG	1377
Db	1329	TTACAATPAGGCTTATGACTTACTAGCAAGATTTACCAAGATTTACTTGATAATAAAG	1388
QY	1378	TCGTAATTCGATTTCCOAGCCTTAGACAAATTTATTAGAACGCTTGAAATGATGAATCGAC	1437
Db	1389	TCGACAAGTTGATTTTCAGGCTTTGGTAACCTCTTGGAAACGACTCAAGGATGCTCAAG	1448
QY	1438	TAATAAAGAAAAATTTGGTAPAGATGATTTATTGGCATCTCTAGCACAATTAACCATCCAGA	1497
Db	1449	TGATAAGTCAAGTTAGTGGAGATATTTCTTGCCCTTCTTAGCTCCGATTCGTATCCAGA	1508
QY	1498	CGCACTTGGCAAAACCAATTTCTCAAATTTAGTATPATCTGAAGACGAAGTTCGTATGTCTCA	1557
Db	1509	ACGTTTACGAAAAACCAATGCGCAATTAACCTACCTGATGAGATTCAGAGTAGCCAA	1568
QY	1558	ATTAGCTGATAAGTATACAACTGACAGTGTATCAATTTTCATGAAACATCATATAATCAG	1617
Db	1569	GTTCGACGCGAAGTACACAGCAGAGACGGTTATCTTTGATCCTCGTATATATACCAG	1628
QY	1618	TGATCAAGGAGATGCATATGTAAACGCTTCATATGGCCATAGTCACCTGGATTCGAAAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAACTCCACATATGACCATAGGCACCTGGATTTAAAAAGA	1688
QY	1678	TAGCCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATACTATAAGAAAAGTATCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGACGGCGACCCAGGCTTATGCTTAAAGAAAAGGTTTGAC	1748
QY	1738	ACCTCCATCTCCAGACGCGAGATGTTTAAAGCAAACTCCAAGTATAGTCAGCAGCAGCTAT	1797
Db	1749	CCCTCTCTTCGACAGACCATCAGGATTCAGGAAATACTCAGGCCAAAAGCAGAGAAGCTAT	1808
QY	1798	TTACAATCGTGTGAAGGGGAAAAACGAATTTCCACTCGTTGCACTTCCATATATGTGTTGA	1857
Db	1809	CTACAACCGMTGAAAGCAGCTTAAGAAGTGCCACTTGATCGTATGCTTACATCTTCA	1868
QY	1858	GCATACAGTTGAGGTTTAAAAACGGTAAATTTGATTTATCTCATTAAGGATCATTAACCAATA	1917
Db	1869	ATATACCTGTGAAAGTCAAAAAACGGTAGTTTATATCACTCATTTATGACCTTACCAATA	1928

QY	1918	TATTAAATTTGCTTGGTTTGAGTATCA	CATACATACAAAGCTTCCAATCGCTTACCTTTGGA	197
Db	1929	CATCAAAATTTGAGTGGTTTTCAGCAAGCGCC	TTTATATGAGGCACTTAAAGGGGTATATCTTTGA	1988
QY	1978	AGATTCTGTTTGGCAGCAAGTAAAGTACTACG	TAGTAACACCCCTGACGAACGTCCACATTTCTAA	2037
Db	1989	GGATCTCTTTTGGCGACTGTCTAAGTACTATGT	CGACATCCAAACGACGTCCGCATTCAGA	2048
QY	2038	TGATGATGGGGCAATCCAGTGTAGCATGTGT	TAGGCAAGAAAGACACACAGTGAAGATCC	2097
Db	2049	TAATGGTTTGGTTAAACGCTAGCGACCATGTT	CAAAAGAAACAAAATATGCTCAAGCTGATAC	2108
QY	2098	AAATAAGAACTTCAAGCGGATGA	-----	2121
Db	2109	CAATCAAAACGAAAAACCAAGCGGAGGAGAA	ACCTCAGACAGAAACCTGAGGAAGAAAC	2168
QY	2122	-----	-----	2130
Db	2169	CCCTCGAAGAGAGAAACCGCAAAACGAGAA	ACCAGAGTCTCCAAAACCAACAGAGGAACC	2228
QY	2131	AGAGAAACACCTGTGAGCAGAAAGTCCCTCA	AGTAGAGACTGAAAAGTAGAAGCCCA	2190
Db	2229	AGAAAGATCACAGAGGAATCAGAAGAACCTC	CAGGTCGAGACTGAAAAGGTTTGAAGAAA	2288
QY	2191	ACTCAAGAAGCAGAGTTTGTCTTGGCAAAGT	TAACGATTCTAGTCTGAAAGCCCAATGC	2250
Db	2289	ACTGAGAGAGGCTGAAGATTACTTGTGAAA	AAATCCAGATCCAAATATCAAGTCCATGC	2348
QY	2251	AACAGAACTCTAGCTGGTTTACGAAATAAT	TTTGACTCTTCAAAATATGATATACAATAG	2310
Db	2349	CAAGAGACTCTCAGAGTTAAAAATAATTTAC	TATTGGCACCCAGGACAAACAATAC	2408
QY	2311	TATCATGCGAAGCAGAGAAATTTACTTGC	GTTGTTAAAGGAGTAA 2358	
Db	2409	TATTATGCGAAGCTGAAAACATTGGCTTT	TATTAAAGGAGAGTAA 2456	
RESULT 23				
BD268049		BD268049	2531 bp	DNA linear PAT 17-JUL-2000
LOCUS		Streptococcus pneumoniae protein and immunogenic fragments for		
DEFINITION		vaccines.		
ACCESSION		BD268049		
VERSION		BD268049.1	GI:33077817	
KEYWORDS		JP 2002532561-A/7.		
SOURCE		Streptococcus pneumoniae		
ORGANISM		Streptococcus pneumoniae		
		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
REFERENCE		1 (bases 1 to 2531)		
AUTHORS		Johnson, L.S., Koenig, S. and Adamou, J.E.		
TITLE		Streptococcus pneumoniae protein and immunogenic fragments for		
JOURNAL		Patent: JP 2002532561-A 7 02-OCT-2002;		
COMMENT		MEDIMMUNE INC		
		OS Streptococcus pneumoniae		
		PN JP 2002532561-A/7		
		PD 02-OCT-2002		
		PF 21-DEC-1999 JP 2000589215		
		PR 21-DEC-1998 US 60/113048		
		PI LESLIE S. JOHNSON, SCOTT KOENIG, JOHN E. ADAMOU		
		PC AG1K39/09, AG1K39/00, AG1P31/10, AG1K37/02		
		CC Streptococcus pneumoniae protein and immunogenic fragments for		
		CC vaccines		
PH	Key	Location/Qualifiers		
FT	source	1..2531	/organism='Streptococcus pneumoniae'.	
FEATURES		Location/Qualifiers		
source		1..2531	/organism="Streptococcus pneumoniae"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:1313"	

Query Match		41.5%; Score 991; DB 6; Length 2531;
Best Local Similarity		65.7%; Pred. No. 3.2e-201;
Matches 1609, Conservative		1; Mismatches 697; Indels 141; Gaps 6;
Qy	1	TTCTTACGAGTGGGACTGTATCAAGCTAGACGGTTAAGGAAAA---TAATCGTGTTC 57
Db	60	TTCTATGAGCTTGGACGTTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119
Qy	58	CTATATAGATGGAACAAAGCGGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAG 117
Db	120	TTATATAGATGGGTGATCAGGCTGGTCAAAAGGCGAGAAACTTGACACAGATGAAGTCAG 179
Qy	118	CAAGCGTGAAGGAATCAATCTCGAGCAATTCGTCATCAAGATAACAGCAACGAGCTATGT 177
Db	180	TAAGAGGAGGGATCAACGCCGACAAATGTTATCAAGATTACGGATCAAGGTTATGT 239
Qy	178	CACCTCACATGGCGACCATATCATTATTAATGTTAAGTTCCTTTATGACGCTATCAT 237
Db	240	GACCTCTCATGGAGACCATATCATTTACTATATATGCGAAGTTCCITATGATGCCATCAT 299
Qy	238	CAGTGAAGATTTACTCATGAAGATCCAACTATTAAGCTAAAGATGAGGATTTGTTAA 297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAAGTTGAAGGATTCAGACATTTGTC 359
Qy	298	TGAGGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATACTATGTTTACCTTTAAGGA 357
Db	360	TGAATCAAGGGTGGTTATGTCATTAAAGGTAAAGTTAACTATGTTTACCTTTAAGGA 419
Qy	358	TGCTGCCACGCGGATTAACGTCGTTCAAAAGAGGAAATCAATCGACAAAAACAAGACA 417
Db	420	TGCRGCTCATGCGGATAATATTCGGACAAAGAGAGATTTAAAGCTCAGAGCAGGAACG 479
Qy	418	TAGTCAACATCGTGAAGTGGAACTCCAAGAAAGCATGGTCTGTTGCTTGGCATTGTC 477
Db	480	CAGTCAATATCAT-----AACTCAAGAGCAGATTAATGCTGTTGCTGCAGCCAGAGC 530
Qy	478	GCAAGGACGCTATPACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATACAGGA 537
Db	531	CCAAGGACGTTATACACGGATGATGGGTATATCTTCAATGCACTGATCATCATGAGGA 590
Qy	538	TACTGTGTATGCTTATATCGTTCTCATGGAGATCATTAACATTAATTCCTAAGAAATGA 597
Db	591	CACGGGTGATGCTTATATCGTTCTCATCGCGACCATTAACATTAATTCCTAAGAAATGA 650
Qy	598	GTATCAGCTAGCGGTGCTGCTGACAGAGCTTCTTATCTGTCGAGGAAATCTGTC 657
Db	651	GTATCAGCTAGCGGTGCTGCTGACAGAGCTTATGGAATGG----- 695
Qy	658	AAATTCAGAACCTATFCGCGGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACC 717
Db	696	-GAAGCAGGATCTCGTCTTCTTCAAGTTCCTAGTTTAAATGCAATCCAGCTCAACCAA 754
Qy	718	TTCTGTAGCAATCCAGGAACTACAAATTAACACAGCAACAACAGCAACACTAACAG 777
Db	755	GATTTGTCAGAGAACCAATCT-----GACTGTCACTCCAATTA 794
Qy	778	TCAAGCAAGTCAAGATTAATGACATTAATGATCTCTTGAACAGCTCTCAAACTGCCCTT 837
Db	795	TCATCAAAATCAAGGGGAAAAATTTCAAGGCTTTTACGTGAATTTGATGCTTAAACCCCT 854
Qy	838	GAGTCAACGACATGTAGATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAGTCG 897
Db	855	ATCAGAACCCATGTGGATCTGATGGCCTTATTTTCGACCCAGCGCAATCACAGTCG 914
Qy	898	ACAGGTAGAGTGTTCAGTGCACACAGGATCATTAACACTTCACTCCCTTACTCTCA 957
Db	915	AAACCGCAGAGGTAGTGTGCTCATGTTAACCATTAACACTTTATCCCTTTATGAACA 974
Qy	958	ATGTCGAATTCGAAGAACGATCGCTGTTATTTATTCCTTCTGTTATCGTTTCAAAACA 1017
Db	975	ATGTCGAATTCGAAGAACGATTCGTTATTTATTCCTTCTGTTATTCCTTCTGTTTCAAAACA 1034

Qy	1018	TTGGTACCAAGATTCAAGGCCAGAACCAACAGTCCACAAACCGACTCCGGAACCTAGTCC 1077
Db	1035	TTGGTACCAAGATTCAAGGCCAGAACCAAGTCCACAAACCGACTCCAGAACCTAGTCC 1094
Qy	1078	AGGCCGCCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCT 1137
Db	1095	AGTCCGCCAACCAAGCTCCAAAGCAATCCAATTTGATGGGAA-----ATTGGTCAAGGAAGC 1148
Qy	1138	GGTACGAAAGTTGGGAGGATATGTTATTCGAGAAAAAGGCACTCTCTGTTATGTCCT 1197
Db	1149	TGTTCCGAAAGTAGGCGATGTTATGCTTTGAGGAGATGGAGTTCTCGTTATATCCC 1208
Qy	1198	TGCGAAAGATTTACCATCTGAAACTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGA 1257
Db	1209	AGCCAAGGATCTTTACGACAGAAACAGCAGCAGGCAATGATAGCAAACTGCCCAAGCAGGA 1268
Qy	1258	GAGTGTTCACACACTTTAACTGCTTAAAGAAAAATGTTGCTCCTCGTGACCAAGAAAT 1317
Db	1269	AGTGTATCTCATAGCTAGGAACTAAGAAAACTGACCTCCCATCTAGTATCGAAGAT 1328
Qy	1318	TTATGATAAAGCATATAATCTGTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGG 1377
Db	1329	TTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTTGATTAAGG 1388
Qy	1378	TGCTAATCTGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGAC 1437
Db	1389	TCGACAGTGTGATTTTGGAGCTTTGGATAACCTGTTGGAAACGACTCAAGGATGCTCAAG 1448
Qy	1438	TTATAAGAAAAAATTCGTAGATGATTTATTGGCATTTCTAGCACCATTAACCCATCCAGA 1497
Db	1449	TGATAAAGTCAAGTTAGTGGAGATATCTTGCCTTTAGCTCCGATTCGTCATCCAGA 1508
Qy	1498	GGGACTTGGCAACCAAAATTCCTCAATGAGTATCTGAAGACGAAAGTTCGTTATGCTCA 1557
Db	1509	ACGTTTAGGAAAAACCAATTCGCAAAATTTACCTACACTGATGATGAGATTCAGTAGCCAA 1568
Qy	1558	ATTAGCTGATAGTATACAAACGTCAGATGTTACATTTTGGTGAACATGATATAATCAG 1617
Db	1569	GTGGCAGCAAGTACACAGCAGAAACGCTTATCTTTGATCCTCGTGAATAACCCAG 1628
Qy	1618	TGATGAAGGAGATGCAATATGTAACGCTCATATGGGCCATAGTCACTGGATTCGAAAAAGA 1677
Db	1629	TGATGAGGGGATGCTATGTAACCTCACTATGATGATGATGATGATGATGATGATGATGAT 1688
Qy	1678	TAGCCTTTCTGATAAGGAAAAAGTTGACGCTCAAGCCTATCTAAGAAAAAGGATTCCT 1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGAGCGGACGCCAGGCTTATGCTRAAGAGAAAGGTTGAC 1748
Qy	1738	ACCTCCATCTCCAGAGCAGATGTTAAAGCAATCCAACTGGAGATAGTGGAGAGCTAT 1797
Db	1749	CCCTCCTTCGACAGACCAATCAGGATTCAGGAAATATCTGAGGCAAAAGGAGAGAGCTAT 1808
Qy	1798	TTACAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCCGACTTCCATATATGTTGA 1857
Db	1809	CTACAACCGMTGAAAGCAGCTAAGAGGTGCCATTCATGATGCTATGCTTCAATCTTCA 1868
Qy	1858	GATACAGTGTAGGTTTAAACCGTAAATTTGATTAATCTCTCAAGGATTCATTACCATAA 1917
Db	1869	ATATACTGTAGAAGTCAAAAACCGTAGTTTAACTACCTCATTTATGACCAATACCATAA 1928
Qy	1918	TATTAATTTCTGCTGTTTGTATGATCACATACAAAGCTCAAAATGGCTATACCTTGA 1977
Db	1929	CATCAAAATTTGAGTGGTTTGACGAAGGCTTTTATGAGGCACTTAAGGGGTATACTCTTGA 1988
Qy	1978	AGATTGTTTTCGAGCATTAAGTACTAGTGAACACCCCTGACCAACCTCCACATTTCTAA 2037
Db	1989	GGATCTTTTGGGACTGTCAAGTACTATGTCGAACATCCAAACGAAACCTCCGATTCAGA 2048
Qy	2038	TGATGATGGGCAATGCCAGTGAAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCC 2097
Db	2049	TAATGGTTTGGTAACGCTAGCAGCAATGTTCAAGAAACAAAAATGCTCAAGCTGATAC 2108
Qy	2098	AAATTAAGAACTTCAAGCGGATGA----- 2121

Db 2109 CAATCAACGGAACCAACGAGGAGAGAAACCTCAGACAGAAAAACCTGAGGAGAAC 2168
Qy 2122 -----AGAGCCAGT 2130
Db 2169 CCCTCGAAGAGAGAAACCGCAAGCGAGAAACACAGAGTCTCCAAACCAACAGAGGAACC 2228
Qy 2131 AGAGGAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTCAAAAAGTGAAGGCCA 2190
Db 2229 AGAAGATCAACGAGGAATCAGAAGAACTCAGGTCGAGACTGAAAAGGTGAAGAAA 2288
Qy 2191 ACTCAAGAGAGCAAGCTTTGCTTGCAGAAAGTAACGGATTCTAGTCTGAAAGCCAATGC 2250
Db 2289 ACTGAGAGAGCTGAAGATTACTTTGGAAAAATCCAGATCCAAATTATCAAGTCCAATGC 2348
Qy 2251 AACAGAAACTCTAGCTGGTTTACGAATAATTTGACTCTTCAMATTATGATACAAATAG 2310
Db 2349 CAAAGAGACTCTCAGAGATTAAAAAATAATTTACTATTATTGGCACCCAGGACCAACAATC 2408
Qy 2311 TATCATGCGCAGACGACAAAAATTTACTTGGCTTGTAAAAGGAAGTAA 2358
Db 2409 TATTATGGCAGAGCTGAAAAACTATTGGCTTTTAAAGGAGGTAA 2456

RESULT 24
AR344448
LOCUS AR344448 2531 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6582706.
ACCESSION AR344448
VERSION AR344448.1 GI:33740487
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2531)
AUTHORS Johnson,L.S. and Adamou,J.E.
TITLE Vaccine compositions comprising Streptococcus pneumoniae polypeptides having selected structural MOTIFS
JOURNAL Patent: US 6582706-A 11 24-JUN-2003;
FEATURES Location/Qualifiers
1..2531
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 41.5%; Score 991; DB 6; Length 2531;
Best Local Similarity 65.7%; Pred. No. 3.2e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;
Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
Db 60 TTCCTATGAGCTTGGACGTTACCAAGCTGTGTCAGATTAAGAAGAGTCTAATCGAGTTGC 119
Qy 58 CTATATAGATCGAAACAAAGCGACGCCAAAAACCGAGAAATTGACTCTGATGAGTTAG 117
Db 120 TTATATAGATGGTATGATCAGGCTGGTCAAAAGGCGAGAAAACCTTGACACAGATGAAGTCAG 179
Qy 118 CAAGGCTGAAGGAATCANTGTGACCAATCGTCATCAGATTAACAGACCAAGGCTATGT 177
Db 180 TAAGAGGAGGGGATCAACGCGCGAAATTTGTTATCAAGATTACGATCAAGGTTATGT 239
Qy 178 CACTTCACATGGCGACCACTATCATTTATCAATGGTAGGTTCCCTATGACGCTATCAT 237
Db 240 GACCTCTCATGAGACCATTATCATTTATTAATGCAAGGTTCTTTATGATGCCATCAT 299
Qy 238 CAGTGAAGATTATCTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAGTTGAAGGATCAGACATTTGTCAA 359
Qy 298 TGAGGCTCAAGGCTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAAGA 357
Db 360 TGAATCAAGGGTGGTTATGTCATTTAGGTAAACGGTAAATACATGTTTACCTTAAAGA 419

Qy 358 TGCTGCCCGCGGATAACGTCGGTACAAAAGAGGAATAACAATCGACAAAAACAAGAGCA 417
Db 420 TGCRCGCTCATCGGATAAATATTCGGAACAAAGAGAGATTAAACGTCAGAGCAGGAACG 479
Qy 418 TAGTCAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGTTCCTTGGCAGCTTC 477
Db 480 CAGTCAATAATCAT-----AACTCAAGAGCAGATAATGCTGTGCTCGACCGAGGC 530
Qy 478 GCAGGAGCGCTATACAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGA 537
Db 531 CCAGGAGCGTTATACACGATGATGGGTATATCTTCAATGCATCTGATATCATTAGGGA 590
Qy 538 TACTGTGTGATGCTTATATCGTCTCATGAGAGTATTACATTACATTCTCCTTAAGATGA 597
Db 591 CACGGGTGATGCTTATATCGTCTCACGGCGACCATTAACATTACATTCTTAAGATGA 650
Qy 598 GTTATCAGCTAGCGAGTGGCTGCTCAGAGAGCTTCTCTATCTGGTCGAGGAATCTGTC 657
Db 651 GTTATCAGCTAGCGAGTGGCTGCTCAGAGAGCTTCTTGAATGG----- 695
Qy 658 AATTTCAAGAACCTTATCGCGACAAAATAGCGATAACAACCTTCAAGAAACAACTGGGTACC 717
Db 696 -GAGCAGGGATCTCGTCTCTTCTCAAGTTCAGTTATATGCAATCCAGCTCAACCA 754
Qy 718 TTCTGTAAAGCAATCCAGGAACTACAAATATACACAAAGCAACACAGCAACATAACAG 777
Db 755 GATTGTCAAGAACCAACAATCT-----GACTGTCACTCCAACTTA 794
Qy 778 TCAAGCAAGTCAAGTAAATGACATTTAGTCTCTTGAAACAGCTCTTACAACTGCCTTT 837
Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTT 854
Qy 838 GAGTCAACGACATGTAGAATCTGATGGCCCTTCTTTGATCCAGCACAAATFCAAGTCG 897
Db 855 ATCAGAACGCCAATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATFCAAGTCG 914
Qy 898 AACAGCTAGAGGTGTTGGAGTGCACACGAGAGATCATTAACACTTATCCCTTACTCA 957
Db 915 AACCGCAGAGGTGTAGCTGTCTCATGTGTAACCATTAACACTTATCCCTTATGAACA 974
Qy 958 AATGCTGTAATTGGAAGACGAATCGCTCGTATTATTTCCCTTCTGTTATCTTCAACCA 1017
Db 975 AATGCTGTAATTGGAAGACGAATCGCTCGTATTATTTCCCTTCTGTTATCTTCAACCA 1034
Qy 1018 TTGGGTACCAGATTCAAGGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db 1035 TTGGGTACCAGATTCAAGCCAGAACCAAGTCCACACCGACTCCAGAACCTAGTCC 1094
Qy 1078 AGGCCCGCACTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCT 1137
Db 1095 AAGTCCGCAACAGCTCCAAGCAATCCAATTTGATGGGAA-----ATTGGTCAAGAGC 1148
Qy 1138 GGTACGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTCGTTATGCTTT 1197
Db 1149 TGTTCGAAAAGTAGCGATGGTTATGCTTTTGAGGAGATGAGATTCTCGTTATATCCC 1208
Qy 1198 TCGAAAAGATTACCATCTGAAACTGTTTAAAAATCTTGAAGCAAGATTATCAAAAACA 1257
Db 1209 AGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATTTGATAGCAAACTGGCCAGCAG 1268
Qy 1258 GAGTGTTCACACACTTTAACTGTAAAGAAAGAAATGTTGCTCTCTCGTACCAAGAAAT 1317
Db 1269 AAGTTTATCTCATAGCTAGGAACTTAAGAAACATGACCTCCCATCTAGTATCGGAAT 1328
Qy 1318 TTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGG 1377
Db 1329 TTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTTACTTGTATAAAG 1388
Qy 1378 TCGTAATTTCTGATTTCCAGCCCTTAGACAAATATTAGACGCTTGATGATGATCGAC 1437
Db 1389 TCGACAAAGTTGATTTTGGAGGCTTTGGATAACCTGTGGAACGACTCAAGGATGTCCTCA 1448
Qy 1438 TAAATAAGAAAAATGGTAGATGATTTTATTTGGCAATTCCTGACCAATTTACCATCCAG 1497

Db	1449	TGATAAAGTCAAGTTAGTGAAGATAATCTTGGCTTCTTAGCTCCGATTCGTATCCAGA	1508
Qy	1498	GCGACTTGGCAAAACCAAAATTCCTCAAAATAGTATACCTGAAGACCAAGTTGCTATTGCTCA	1557
Db	1509	ACGTTTAGGAAAACCAATCGCAAAATTCCTACACTGATGATGAGATTCAAGTAGCCAA	1568
Qy	1558	ATTAGCTGATAAGTATACAACGTGAGATGGTTACATTTTGTAGAAACATGATATAATCAG	1617
Db	1569	GTTGGCAGGCAAGTACACACAGAGAAGCGGTATATCTTTGATCCTCGTGATATAACCA	1628
Qy	1618	TGATGAAGAGAGATGCATATGTAAACGCTCATATGGCCATAGTCACATGATTCGAARA	1677
Db	1629	TGATGAGGGGATGCCTATGTAACTCCACATATGACCATAGCCACTGGATTAAAAAGA	1688
Qy	1678	TAGCCTTTCTGATAAGGAAAAGTTGCAGTCAAGCCCTATCTAAAGAAAAGGTATCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCAGGCTTATGCTRAAGAGAAAGTTTGAC	1748
Qy	1738	ACCTCCATCTCCAGACGCAGATGTTAAAGCAAAATCAACTGGAGATAGTGCAGCAGTAT	1797
Db	1749	CCCTCCTTCACAGACCAATCAGGATTCAGAAATATCTGAGCAAAAGGACGAAAGTAT	1808
Qy	1798	TTACAAATCGTGTGAAAGGGGAAAACAAATTCCACTCGTTCGACTTCCATATATGTTGA	1857
Db	1809	CTACAAACCGMGTGAAAGCAGCTAAGAGAGTGCCACTTGATCGTATGCCTTACAATCTCA	1868
Qy	1858	GCATACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACATAA	1917
Db	1869	ATACTGTTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTTATGACCAATTACCATA	1928
Qy	1918	TATTAATTTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGA	1977
Db	1929	CATCAAAATTTGAGTGGTTTCACGAGGCCCTTTATGAGCACCTTAGGGTATACTCTGA	1988
Qy	1978	AGATTTGTTGACAGATTAAGTACTAGTAGAACACCTTGACGAACGTCCACATTTCAA	2037
Db	1989	GGATCTTTTGGCCACTGTCAAGTACTATGTGCAACATCCAAAGCAACGTCCGCAATCAGA	2048
Qy	2038	TGATGATGGGGCAATGCCAGTCAGCATGTGTTAGCAGAGAAGACCAACAGTGAAGTCC	2097
Db	2049	TAAATGGTTTGGTTAAACGCTAGCCACCATGTTCAAGAAACAAAATGGTCAAGCTGATC	2108
Qy	2098	AAATAAGAACTTCAAAGCGGATGA-----	2121
Db	2109	CAATCAAAACGAAAAACCAAGCAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAC	2168
Qy	2122	-----AGAGCCAGT	2130
Db	2169	CCCTCGAAGAGAAACCGAAACGAGAGAAACGAGTCTCCAAAACCAACAGAGAAC	2228
Qy	2131	AGAGGAAACCTTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCA	2190
Db	2229	AGAAGAATCACAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAGAGTTGAAGAAA	2288
Qy	2191	ACTCAAGAGACAGAGTTTGTCTTCGGAAGTAAACGGATTCTAGTCTGAAGCCAAATGC	2250
Db	2289	ACTGAGAGAGGCTGAAGATTACTTGGAAAAATCCAGGATCCAAATTTATCAAGTCCAATGC	2348
Qy	2251	AACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAATTTATGATAACAATAG	2310
Db	2349	CAAGAGACTCTCACAGGATTAAGAAATTAATTTACTATTGGCACCCAGACCAACATAC	2408
Qy	2311	TATCATGGCAGAAGCAGAAAAATTACTTGGTGGTTTAAAGGAGTAA	2358
Db	2409	TATTATGGCAGAGCTGAAAACTATTGGCTTTATTAAAGGACAGTAA	2456

RESULT 25	AF318955	LOCUS	AF318955	2517 bp	DNA	linear	BCT 11-FEB-2001
DEFINITION	Streptococcus pneumoniae pneumococcal histidine triad protein D precursor (phtD) gene, partial cds.						

ACCESSION AF318955 GI:12744743
 VERSION AF318955.1
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 2517)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T.,
 Dormitzer, M., Dagan, R., Brewah, Y.A., Barren, P., Lathigra, R.,
 Langermann, S., Koenig, S. and Johnson, S.
 Identification and characterization of a novel family of
 pneumococcal proteins that are protective against sepsis
 Infect. Immun. 69 (2), 949-958 (2001)
 21101045
 PUBMED 11159990
 2 (bases 1 to 2517)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M. and
 Johnson, S.
 Direct Submission
 Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
 West Watkins Mill Road, Gaithersburg, MD 20878, USA
 Location/Qualifiers
 1..2517
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="4"
 /db_xref="taxon:1313"
 1..>2517
 /genes="phtD"
 1..>2517
 /gene="phtD"
 /note="PhtD; protective pneumococcal antigen"
 /codon_start=1
 /transl_table=11
 /product="pneumococcal histidine triad protein D
 precursor"
 /protein_id="AAK06760.1"
 /db_xref="GI:12744744"
 /translation="MKINKYLAGSVAVLALSVCSELGRHQGVQKESNRVSVDIIE
 DQAGKAENLPDEVSREGINAEQIVIKITDQYVTSHGDIHYNGKVPDYDAIIE
 ELLMDPNYQLKDSIVNKGIVKDGQYVYLKDAHADNITKEEIKRQKEH
 SNHGGSDNVAARAAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHVPKNE
 LSASLAAEAAYWKGQGRSPSSSYNANPAQRLSENHNLTVPTPTHQNGENISS
 LARELVPLSRHVESDGLIFDPAQITSETARGVAVPHGNHYHFTPYEQMSELEKRI
 ARILPLRVSNHWPDSRPEQSPQSTPEPSPQAPNPQAPSNFIDKLVKEAVR
 KYGDGVFEENGVSRIIPAKDLSAETAAGIDSKLAKQESLSHKLGAACKTDLPSSREF
 YNKAYDLARLHQDLDNKGQVDFEALNDLLERLKDVPSDKVLGVDVDDILAFIAPIRH
 PERLQKNPAQITYTDETQVAKLAGYTTEDGYIFDPRDITDEGDAYVTPMTHGSH
 IKKDSLPAERAAQAAYAKEKGLTPPTSDQSGNTEAKGARAIYNRVKAACKVPLDR
 MPNYLQYTVKNGSLIIPHYDHYNIKFEWDEGLYEAPKQYTTLEDLLATVKYVYEH
 PNERPHSDNGFCNASDHYRKNKVDQSDKPEDKEHDEVESEPHSPSDEKENHAGLNPS
 ADNLKYPSDITDEBEERADITDEAIPQVENSVINAKIADAELKVTDPDSIRQNA
 METLTGLKSSLLLTGKDNNTTISAEDVSLALLKESQAPAIQ"
 1..87
 /gene="phtD"
 247..264
 /gene="phtD"
 /note="Region: histidine triad"
 418..495
 /gene="phtD"
 /note="Region: coiled-coil domain"
 619..636
 /gene="phtD"
 /note="Region: histidine triad"
 943..960
 /gene="phtD"
 /note="Region: histidine triad"
 1417..1509
 /gene="phtD"
 /note="Region: coiled-coil domain"
 1678..1695
 /gene="phtD"

misc_feature		/note="Region: histidine triad"	
		1930..1947	
		/gene="phnD"	
		/note="Region: histidine triad"	
		526..717	
repeat_region		/rpt_type=direct	
repeat_region		1597..1779	
		/rpt_type=direct	
ORIGIN			
Query Match			
Best Local Similarity 68.18; Score 990; DB 1; Length 2517;			
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;			
Qy	1	TTCTTAGAGTTGGGACTGATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC	57
Db	60	TTCTATAGAACTTGGTCTGCTCACCAGCTGCTGCGTTAAGAAAGAGTCTAATCGAGTTC	119
Qy	58	CTATATAGATGGAAAAACAAGCGACGCAAAAAACGGAGAAATTGACTCCTGATGAGTTAG	117
Db	120	TTATATAGATGGTATCAGGCTGCTCAAAAGGCAGAAAACTTGACACCAAGATGAAGTCAG	179
Qy	118	CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAAACAGACCAAGGCTATGT	177
Db	180	TAAAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT	239
Qy	178	CAGTTACATGGCGACCACTATCATTAATTAACAATGGTAAGGTTCTTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGCAATTAATCAATTAATAAGGCAAGGTCCTTATGATGCCATCAT	299
Qy	238	CAGTGAAGATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTGTAA	297
Db	300	CAGTGAAGACTCTCATGAAGATCCGAATTATCAGTTGAAGATTACAGACATTGCA	359
Qy	298	TCAGGTCAAGGTTGGATATGTTATCAAGTATAGTGAAGAAATACATATGTTTACCTTAGGA	357
Db	360	TGAATCAAGGCTGGTTATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGA	419
Qy	358	TGCTGCCACCGGATACGTCCTGACAAAGAGGAATCAATCGACAAAAACAAGAGCA	417
Db	420	TGCAGCTCATCGGATATATTCGGACAAAGAGAGATTAAACGTGAGAGCAGGAAACA	479
Qy	418	TAGTCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTCTGTGCTTGGCACGTTTC	477
Db	480	CAGTCAATCAACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGTCAGCCAGAGC	533
Qy	478	GCAAGGACGCTATACCTACAGATATGTTATCTTTAATGCTCTGATATCATAGAGGA	537
Db	534	CCAAAGGACGCTATACAACGGATGATGGTTATATCTTCAATCTGATATCATTTGAGGA	593
Qy	538	TACTGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAACATTCTTAAGAAATGA	597
Db	594	CACGGGTGATGCTTATATCGTTCCTCAGCGGACCAATTAACCATTAACATTCTTAAGAAATGA	653
Qy	598	GTTATCAGCTAGCAGTGGTGTGCTGCAAGAGCCCTTCCTATCTGGTCGAGGAAATCTGTC	657
Db	654	GTTATCAGCTAGCAGTGGTGTGCTGCAAGAGCCCT-----	688
Qy	658	AAATTCAGAACCTTATCCCGGACAAATATAGCGATTAACCTCAAGAAACAACTGGGTACC	717
Db	689	-----ATTGGAAATGGGAAGCAGGGATCTGCTCTTTCTTTCAAGTTCTAGTTATAATGC	740
Qy	718	TTCTGAAGCAATCCAGAACTACAAATATCAACAAAGCAACACAGCAACACTTAACAG	777
Db	741	AAATCCAGCTCAACCAAGATTGTCAGAGNACCAGAAATCTGACTGTCACTCCCACTTATCA	800
Qy	778	TCAAGCAAGTCAAAGTAATGACATTTGATGTCCTTTGAAAACAGCTCTCAAACTGCCCTTT	837
Db	801	TCA---AAATCAAGGGGAAAAATTCAGAGCCCTTTTACGTGAATTTGATGCTTAACCCCTT	857
Qy	838	GAGTCAACGACATGTAGATCTGATGGCGCTGCTCTTGTATCCAGCACAAATCACAAGTCG	897
Db	858	ATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAATCACAAGTCG	917

Qy	898	AACAGCTAGAGGTGTTGCAGTGGCCACACGAGATCAATTACCACCTTCATCCCTTACTCTCA	957
Db	918	AACCGCCAGAGGTGTAGTGTCTCATGTGTAACTTACCACCTTATCCCTTATGAACA	977
Qy	958	AATGCTGAATTGGAAGACGAATCGCTCGTATATATCCCTTCGTTATGTTTCAACCA	1017
Db	978	AATGCTGAATTGGAAGAACGAATGCTCGTATATATCCCTTCGTTATGTTTCAACCA	1037
Qy	1018	TTGGGTACAGATTTCAGGCGCAGAAACAACCAAGTCCACACCGCACTCCGAACTAGTCC	1077
Db	1038	TTGGGTACAGATTTCAGGCGCAGAAACAACCAAGTCCACACCGCACTCCGAACTAGTCC	1097
Qy	1078	AGCCCGCAACCTGCACCAAACTCTTAAATATAGACTCAAAATCTTC-----TTT	1125
Db	1098	AAGTCCGCAACCTGCACCAAACTCTTAAATATAGACTCAAAATCTTC-----TTT	1157
Qy	1126	GTTTAGTCAGCTGGTACGAAAGTGGGGAAGATATGTTATCGAAGAAAGGCACTC	1185
Db	1158	GGTCAAGAAAGCTGTTTCGAAAGTAGGCGATGTTATGTTCTTTGAGGAGATGAGTTTC	1217
Qy	1186	TCGTTATGCTTTTCGAAAGATTTACCATCTGAAACTGTTTAAAAATCTTGAAACCAAGTT	1245
Db	1218	TCGTTATATCCAGCCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAACT	1277
Qy	1246	ATCAAAAACAAGAGAGTGTTCACACACTTTAACTGTCTTAAAGAAATAATGTTGCTCTCG	1305
Db	1278	GGCCCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCCATCTAG	1337
Qy	1306	TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGT	1365
Db	1338	TGATCGAAGATTTTACAAATGAGGCTTATGACTTACTAGCAAGAAATCACCAAGATTTACT	1397
Qy	1366	TGNAATTAAGGCTCGTAAATTTCTGATTTCCAAAGCTTAGACAAATATTAGAACCTTGAA	1425
Db	1398	TGATAATTAAGTTCGACAGATTTGATTTTGAGGCTTTGAGATAACCTGTTGGAACGACTCA	1457
Qy	1426	TGATGAATTCGACTAATTAAGAAATAATTTGATAGATGATTTATTTGGCATCTCTAGCACCAT	1485
Db	1458	GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGTCCGAT	1517
Qy	1486	TACCATTCGAGCGACTTCGCAACCAAAATTCCTCAAAATGAGTATCTAGAAACGAGAT	1545
Db	1518	TCGTCATCCAGACGTTTAGGAAACCAAAATTCGCCAAATATACCTACACTGATGATGAT	1577
Qy	1546	TCGTTATGCTCAATTAGCTGATAAGTATACAACGTCAGATGGTTACATTTTGTGTAACA	1605
Db	1578	TCAAGTAGCCAAAGTTGGCAGCAGTACACAAACAGAGACGGTTATATCTTTGATCCTCG	1637
Qy	1606	TGATAATTAATCAGTGAAGGAGATGCAATATGTAAACGCTCATATGGGCCATAGTCACTG	1665
Db	1638	TGATAATTAATCAGTGAAGGAGATGCAATATGTAAACGCTCATATGGGCCATAGTCACTG	1697
Qy	1666	GATTGGAAAGATAGCCCTTTCTGATAAGGAAAGTTGACGCTCAAGCTTACTATAAAGA	1725
Db	1698	GATTAAAAAGATAGTTTGTCTGAAAGTGAAGAGCGGACGCCAGGCTTATGTAAGA	1757
Qy	1726	AAAAAGTATCTTACCTCCATCTCCAGACGCGAGATGTTTAAAGCAAAATCAACTGGAGATAG	1785
Db	1758	GAAAGTTTACCCCTCTCTCGACAGACCATCAGGATTTCAGGAAATATCTGAGGCAAAAGG	1817
Qy	1786	TGCAGAGCTATTTACAAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGACTTCC	1845
Db	1818	AGCAGAGCTATCTACAAACCGCTGAAAGCAGCTTAAGAAAGTGGCAGCTTGATCGTATGCC	1877
Qy	1846	ATATATGTTTCAGGATACAGTTGAGTTTAAAAACGTTAAATTTGATTTATTCCTCATAGA	1905
Db	1878	TTCAATCTTCAATATCTGTAGAGTCAAAAAACGTTAGTTTATCATCACTCATATGA	1937
Qy	1906	TCATTACCAATATTTAAATTTGCTTGTGTTGATGATCACACATACAAGCTCCAAATGG	1965
Db	1938	CCATTACCAATCAATCAAAATTTGAGTGGTTTGAACGAAGGCCCTTTATAGGCACCTTAAGG	1997

QY	1966	CTATACCTTGGAGATTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACG	2025
Db	1998	GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG	2057
QY	2026	TCCACATTCTAATGATGGATGGGGCAATGCCAGTGCAGTGCATGTGTTAGGCAAGAAAGACCA	2085
Db	2058	TCCGCATTTCAGATAAATGGTTTGGTAACGCTAGCGACCATGTTTCGTAATAAATAAGGTAGA	2117
QY	2086	CAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC	2145
Db	2118	CCAAAGACAGTAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAAGCAACTCACCC	2177
QY	2146	TGAGCCAGAAG	2156
Db	2178	TGAATCTGATG	2188

Search completed: September 30, 2004, 18:25:56
Job time : 9320 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 11:00:45 ; Search time 915 Seconds

(without alignments)

11091.743 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	2	AAV27351 Streptococcus
2	2388	100.0	2389	6	ABQ84819 S. pneumoniae
3	2388	100.0	2389	9	ADC45136 Streptococcus
4	2388	100.0	2406	7	ABX06886 S. pneumoniae
5	2388	100.0	2451	3	AAA47604 Recombinant
6	2388	100.0	8195	2	AAV52227 Streptococcus
7	2388	100.0	110000	7	ABSS56454_09 Continuation (10 o
8	2388	100.0	110000	7	ABSS56454_10 Continuation (11 o
9	1374.2	57.5	2523	3	AAA65731 Streptococcus
10	1374.2	57.5	2647	3	AAA65736 Streptococcus
11	1374.2	57.5	2647	6	ABK15103 DNA encod
12	1003.8	42.0	2478	3	AAA08557 S. pneumoniae
13	993	41.6	2457	3	AAA05417 Streptococcus
14	993	41.5	2531	3	AAA47605 Recombinant
15	990	41.4	2517	7	ABX06705 Streptococcus
16	990	41.4	2531	3	AAA47602 Recombinant
17	990	41.4	2531	3	AAA47602 Recombinant
18	990	41.4	110000	7	ABSS56454_08 Continuation (9 of
19	987.6	41.3	2290	6	AAV27356 Streptococcus
20	987.6	41.3	2290	6	ABQ84824 S. pneumoniae
21	987.6	41.3	2290	9	ADC45146 S. pneumoniae
22	980.8	41.1	2639	3	AAA65737 Streptococcus
23	980.8	41.1	2639	6	ABK15104 DNA encod

24	859.4	36.4	2163	2	AAV25394 Streptococcus
25	653.8	27.4	2359	2	AAV52376 Streptococcus
26	548.4	23.0	3171	3	AAA65739 Streptococcus
27	385.4	16.1	1342	2	AAV27414 Streptococcus
28	385.4	16.1	1342	6	ABQ84882 S. pneumoniae
29	385.4	16.1	1342	9	ADC45262 Streptococcus
30	385.4	16.1	1455	3	AAA05473 Streptococcus
31	385.4	16.1	1455	3	AAZ91804 Streptococcus
32	385.4	16.1	1455	3	AAA65733 Streptococcus
33	385.4	16.1	1455	3	AAA47603 Recombinant
34	385.4	16.1	3117	7	ABX06706 Streptococcus
35	385.4	16.1	3120	3	AAA65730 Streptococcus
36	385.4	16.1	5048	3	AAA65735 Streptococcus
37	385.4	16.1	5048	6	ABK15101 DNA encod
38	385.4	16.1	6867	2	AAV52325 Streptococcus
39	381.4	16.0	973	2	AAV52488 Streptococcus
40	347	14.5	2528	6	AAA65738 Streptococcus
41	347	14.5	2528	6	ABK15105 DNA encod
42	247.4	10.4	504	3	AAA08556 S. pneumoniae
43	243.4	10.2	1684	2	AAV52391 Streptococcus
44	214.6	9.0	2475	6	ABN66838 Streptococcus
45	214.6	9.0	2478	5	AAAS00036 Streptococcus
46	211.4	8.8	1146	3	AAA05814 Group B S
47	211.4	8.8	2466	6	ABN69535 Streptococcus
48	211.4	8.8	2466	6	ABN70334 Streptococcus
49	211.4	8.8	2469	3	AAA05811 Group B S
50	211.4	8.8	2469	5	AAAS00038 Streptococcus
51	211.4	8.8	110000	6	ABN71527_12 Continuation (13 o
52	209.8	8.8	2469	3	AAA65740 Streptococcus
53	209.8	8.8	5215	2	AAV91105 Group B S
54	206.6	8.6	2472	3	AAA65741 Streptococcus
55	206	8.6	110000	6	ABN71527_18 Continuation (19 o
56	201.6	8.4	1398	8	ABN66839 Streptococcus
57	201.2	8.4	816	8	ACF42878 S. pneumoniae
58	201.2	8.4	816	8	ACF42879 S. pneumoniae
59	201.2	8.4	819	6	ABK15106 DNA encod
60	201.2	8.4	819	8	ACF42881 S. pneumoniae
61	200.8	8.4	2721	8	ACF42880 S. pneumoniae
62	188.2	7.9	492	3	AAA05393 Streptococcus
63	117.4	4.9	1455	3	AAA05838 Group B S
64	109	4.6	841	2	AAT28529 S. pneumoniae
65	109	4.6	841	4	ABA76858 Streptococcus
66	102.6	4.3	1137	6	ABN66837 Streptococcus
67	62.8	2.6	2000	7	ADA71938 Rice gene
68	53.2	2.2	1729	2	AAV42952 Streptococcus
69	52.8	2.2	1168	9	ADC87372 Human GPC
70	51.8	2.2	567	7	ABX06707 S. pneumoniae
71	51.8	2.2	1083	5	AAV576745 DNA encod
72	51.6	2.2	3060	4	AAH54131 S. epider
73	51.6	2.2	3309	4	AAH54131 S. epider
74	51.6	2.2	11076	7	ACA47181 Prokaryot
75	51.6	2.2	11091	6	ABN92780 Staphyloc
76	51.2	2.1	15935	4	ABL06114 Streptococcus
77	50	2.1	7662	6	ABN80080 Human cha
78	49.8	2.1	1959	4	AAI14146 Probe #40
79	49.8	2.1	1959	4	ABA55872 Human foe
80	49.8	2.1	1959	4	AAI35531 Probe #42
81	49.8	2.1	1959	4	ABA45382 Human bre
82	49.8	2.1	1959	4	ABA25546 Probe #40
83	49.8	2.1	1959	4	AAK29571 Human bon
84	49.8	2.1	1959	4	AAK04090 Human bra
85	49.8	2.1	1959	4	ABS29202 Human liv
86	49.8	2.1	1959	5	AAI03994 Probe #39
87	49.8	2.1	1959	6	ABS04127 Human gen
88	48.6	2.0	3399	2	AAT05868 Chicken 1
89	48.2	2.0	3579	3	AAV70099 Plasmodu
90	48	2.0	2631	6	ABN68134 Streptococcus

ALIGNMENTS

RESULT 1

AAV27351	DB	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGATAACAGACCAAGGCTATGTCAC	180
ID AAV27351 standard; DNA; 2389 BP.	QY	181	TTCCATGCGCAGCCACTATCAATTAATGTAAGGTTCTTTATGACGCTATCATCAG	240
AC AAV27351;	DB	181	TTCCATGCGCAGCCACTATCAATTAATGTAAGGTTCTTTATGACGCTATCATCAG	240
DT 02-OCT-1998 (first entry)	QY	241	TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTATAAGCTATAAGATGAGGATATGTTAATGA	300
DE Streptococcus pneumoniae SP0036 nucleotide.	DB	241	TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTATAAGCTATAAGATGAGGATATGTTAATGA	300
XX Streptococcus pneumoniae;	QY	301	GCTCAAGGTTGGATATGTTATCAAGGTAGTGAAGAAATCAATCGACAAACAAAGAGCATAG	360
KW detection; pneumonia; otitis media; meningitis; ss.	DB	301	GCTCAAGGTTGGATATGTTATCAAGGTAGTGAAGAAATCAATCGACAAACAAAGAGCATAG	360
XX Streptococcus pneumoniae.	QY	361	TGCCACGCGGATAACGTCGTATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
XX Key	DB	361	TGCCACGCGGATAACGTCGTATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
FH 2. 2389	QY	421	TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTCTTGTGCTTGGACGCTTCGCA	480
FT /*tag= a	DB	421	TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTCTTGTGCTTGGACGCTTCGCA	480
FT /product= "SP0036"	QY	481	AGGACGCTATACAGATGATGTTTATATCTTAAATGCTTCTCATATCATAGAGGATAC	540
FT /transl_except= (pos.1367..1369,aa:Xaa)	DB	481	AGGACGCTATACAGATGATGTTTATATCTTAAATGCTTCTCATATCATAGAGGATAC	540
FT /note= "no stop codon given; Xaa is unspecified"	QY	541	TGGTATGCTTATATCGTCTCTCATGAGATCAATACCATTTACATTCCTAAGAAATGATT	600
XX WO9818930-A2.	DB	541	TGGTATGCTTATATCGTCTCTCATGAGATCAATACCATTTACATTCCTAAGAAATGATT	600
XX 07-MAY-1998.	QY	601	ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTCGAGAAATCTGTCAA	660
XX 30-OCT-1997; 97WO-US019422.	DB	601	ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTCGAGAAATCTGTCAA	660
XX 31-OCT-1996; 96US-0029960P.	QY	661	TTCAAGAACCTATCGCCGACAAATAGCGATTAACATTTCAAGAACAACTCGGTACCTTC	720
XX (HUMA-) HUMAN GENOME SCI INC.	DB	661	TTCAAGAACCTATCGCCGACAAATAGCGATTAACATTTCAAGAACAACTCGGTACCTTC	720
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;	QY	721	TGTAAGCAATCCAGGAATCAAAATACATAACAGCAACCAAGCAACCAAGCAACCTAACAGTCA	780
XX WPI; 1998-272224/24.	DB	721	TGTAAGCAATCCAGGAATCAAAATACATAACAGCAACCAAGCAACCAAGCAACCTAACAGTCA	780
XX P-PSDB; AAW55090.	QY	781	AGCAAGTCAAAAGTAAATGATGATGCTCTTGAACAGCTCTTACAACTGCGCTTCAG	840
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae	DB	781	AGCAAGTCAAAAGTAAATGATGATGCTCTTGAACAGCTCTTACAACTGCGCTTCAG	840
XX - or their epitope-containing fragments, useful in protective or	QY	841	TCAACGATCATGATGATGCTCTTGAACAGCTCTTACAACTGCGCTTCAG	900
XX therapeutic vaccines, and for diagnosis.	DB	841	TCAACGATCATGATGATGCTCTTGAACAGCTCTTACAACTGCGCTTCAG	900
XX Claim 1; Page 59; 118pp; English.	QY	901	AGCTAGAGTGTGTCAGTCCACAGGAGATCAATACCATTTCTTACTCTCAAT	960
XX The present sequence encodes a protein from Streptococcus pneumoniae. The	DB	901	AGCTAGAGTGTGTCAGTCCACAGGAGATCAATACCATTTCTATCCCTTACTCTCAAT	960
XX nucleic acid sequence encoding the Streptococcus pneumoniae protein can	QY	961	GTCGAAATGGAAGAACGAATCGCTGATTAATTTCCCTTCTGTTATCGTTCAAACTG	1020
XX be useful in vaccines for inducing protective antibodies against	DB	961	GTCGAAATGGAAGAACGAATCGCTGATTAATTTCCCTTCTGTTATCGTTCAAACTG	1020
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.	QY	1021	GCTACCATTAAGCCAGAACCAACCAAGTCCACCGACTCCCGAACCTAGTCCAGG	1080
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid	DB	1021	GCTACCATTAAGCCAGAACCAACCAAGTCCACCGACTCCCGAACCTAGTCCAGG	1080
XX are used to detect Streptococcus infection (by usual hybridisation or	QY	1081	CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT	1140
XX amplification methods), also for isolating Streptococcus genes or their	DB	1081	CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT	1140
XX allelic variants. The protein can be used similarly to detect specific	QY	1141	AGGAAAGTTGGGAGGATATGATTCGAGAAAGGAGGACTCTCTCGTTATGCTTTTCG	1200
XX antibodies in standard immunoassays, especially for diagnosing or	DB	1141	AGGAAAGTTGGGAGGATATGATTCGAGAAAGGAGGACTCTCTCGTTATGCTTTTCG	1200
XX monitoring infections. Antibodies which bind the protein are used to	QY	1201	GAAGATTTTACCCTGTAACCTGTTTAAATCTTGAAGCAAGTATCAAAACCAAGAG	1260
XX detect corresponding antigens, to purify the protein and for passive	DB	1201	GAAGATTTTACCCTGTAACCTGTTTAAATCTTGAAGCAAGTATCAAAACCAAGAG	1260
XX immunisation (optionally coupled to a toxin). Vaccines are administered,	QY			
XX e.g. by injection, orally or through the skin, typically at 0.01-1000	DB			
XX (especially 10-300) mu g/ml per dose	QY			
XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;	DB			
XX Query Match	QY			
XX Best Local Similarity 100.0%; Score 2388; DB 2; Length 2389;	DB			
XX Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	DB			
XX 1 TTCTTACGATGGGACTGTATCAAGCTAGACGGTTAAGGAATAATCGTTTCCTA 60	QY			
XX 1 TTCTTACGATGGGACTGTATCAAGCTAGACGGTTAAGGAATAATCGTTTCCTA 60	DB			
XX 61 TATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120	QY			
XX 61 TATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120	DB			
XX 121 GCCTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGATAACAGACCAAGGCTATGTCAC 180	QY			

1261 TGTTTCCACACATTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
 1261 TGTTTCCACACATTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
 1321 TGATAAGCATATAAATCTGTTAACTGAGGCTCATAAAGGCTTTGTTGNAATAAAGGCTCG 1380
 1321 TGATAAGCATATAAATCTGTTAACTGAGGCTCATAAAGGCTTTGTTGNAATAAAGGCTCG 1380
 1381 TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTGAATGATGAATCGACTAA 1440
 1381 TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTGAATGATGAATCGACTAA 1440
 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCGAGCG 1500
 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCGAGCG 1500
 1501 ACTTGGCAAAACCAATTTCTCAAATTTGAGTATCTGAAGACGAAATTCGATTGCTCAATT 1560
 1501 ACTTGGCAAAACCAATTTCTCAAATTTGAGTATCTGAAGACGAAATTCGATTGCTCAATT 1560
 1561 AGCTGATAGTATACAACTGATGATTTACATTTTGTGATGAACATGATATATCATGTA 1620
 1561 AGCTGATAGTATACAACTGATGATTTACATTTTGTGATGAACATGATATATCATGTA 1620
 1621 TGAAGGAGATGATATGTAACGCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680
 1621 TGAAGGAGATGATATGTAACGCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680
 1681 CTTTCTGATAGGAAAAAGTTGCGAGCTCAAGCCTTATCTAAAGAAAAAGGTATCTTACC 1740
 1681 CTTTCTGATAGGAAAAAGTTGCGAGCTCAAGCCTTATCTAAAGAAAAAGGTATCTTACC 1740
 1741 TCCATCTCCAGACGAGATGTTAAGCAAAATCCAACTGGAGATAGTGGAGAGCTATTATA 1800
 1741 TCCATCTCCAGACGAGATGTTAAGCAAAATCCAACTGGAGATAGTGGAGAGCTATTATA 1800
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCGACTTCATATATGTTGAGCA 1860
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCGACTTCATATATGTTGAGCA 1860
 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTTACATAATAT 1920
 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTTACATAATAT 1920
 1921 TAAATTTCTGTTGTTGATGATCACATACAACTGCTCAATGCTATACCTTGGAGA 1980
 1921 TAAATTTCTGTTGTTGATGATCACATACAACTGCTCAATGCTATACCTTGGAGA 1980
 1981 TTTGTTTTCGACGATTAAGTACTAGTATAGAACACCTCGACCAACCTCCACATTTCTAATGA 2040
 1981 TTTGTTTTCGACGATTAAGTACTAGTATAGAACACCTCGACCAACCTCCACATTTCTAATGA 2040
 2041 TGGATGGGGCAATGCGATGAGATGTTTGGAGGAAGAACACCACTGAGATGCCAAA 2100
 2041 TGGATGGGGCAATGCGATGAGATGTTTGGAGGAAGAACACCACTGAGATGCCAAA 2100
 2101 TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC 2160
 2101 TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC 2160
 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAACGAGATTTTGTTCGCGAA 2220
 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAACGAGATTTTGTTCGCGAA 2220
 2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAAAATCTAGCTGTTTACGAAATAA 2280
 2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAAAATCTAGCTGTTTACGAAATAA 2280
 2281 TTTGACTCTTCAATTTATGATTAACATAGTATCATGSCAGAGCAAGAAAATTTACTTGC 2340
 2281 TTTGACTCTTCAATTTATGATTAACATAGTATCATGSCAGAGCAAGAAAATTTACTTGC 2340

2341 GTTGTAAAGGAAGTAATCCTTCATCTCTTAAGTAAGGAAAAAATAAAC 2389
 2341 GTTGTAAAGGAAGTAATCCTTCATCTCTTAAGTAAGGAAAAAATAAAC 2389
 RESULT 2
 ABQ84819
 ID ABQ84819 standard; DNA; 2389 BP.
 XX ABQ84819;
 AC AC
 XX DT 04-SEP-2002 (first entry)
 XX S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW Streptococcus pneumoniae; Streptococcal infection; detection; gene; ds.
 XX Streptococcus pneumoniae.
 OS
 XX US2002061545-A1.
 PN 23-MAY-2002.
 PD 22-JAN-2001; 2001US-00765272.
 XX 30-OCT-1997; 97US-00961083.
 PR (CHOI/) KUNSC H.
 PA (KUNS/) KUNSC H.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 XX (ROSE/) ROSEN C A.
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2002-479261/51.
 DR P-PSDB; ABP54584.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX Claim 1; Page 27; 70pp; English.
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX
 SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;
 Query Match 100.0%; Score 2389; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 TTCTTACAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAAAATATCGTTTCTTA 60
 1 TTCTTACAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAAAATATCGTTTCTTA 60
 61 TATAGATGGAACCAACGACGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 61 TATAGATGGAACCAACGACGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGTAACAGACCAAGCTATGTCAC 180

Db 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACACAGCAAGGCTATGTCAC 180
QY 181 TTCAATGCGCGACCACTATCATTTATTAATGCTAAGGTTCCCTTATGACGCTATCATCAG 240
Db 181 TTCAATGCGCGACCACTATCATTTATTAATGCTAAGGTTCCCTTATGACGCTATCATCAG 240
QY 241 TGAAGAAATTAATCATGAAAGATCCAACTATTAAGCTTAAAGATGAGGATATGTTAATGA 300
Db 241 TGAAGAAATTAATCATGAAAGATCCAACTATTAAGCTTAAAGATGAGGATATGTTAATGA 300
QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAAGGATGC 360
Db 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAAGGATGC 360
QY 361 TGCCACGCGGATTAACGTCGCTACAAAGAGGAAATCAATPCGACAAAACAGAGCATAG 420
Db 361 TGCCACGCGGATTAACGTCGCTACAAAGAGGAAATCAATPCGACAAAACAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGGAATCCAGAAACGATGCTGCTTGGCCTTGGCAGTTCGCA 480
Db 421 TCAACATCGTGAAGGTGGAATCCAGAAACGATGCTGCTTGGCCTTGGCAGTTCGCA 480
QY 481 AGGACGCTATACATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 481 AGGACGCTATACATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY 541 TGGTGATGCTTATATCGTTCCTCATGAGATCAATTACCATTAATTCCTAAGAAATGAGTT 600
Db 541 TGGTGATGCTTATATCGTTCCTCATGAGATCAATTACCATTAATTCCTAAGAAATGAGTT 600
QY 601 ATCAGCTAGGAGTTGCTGCTGAGAGGCTTCTTATCTGCTGAGGAAATCTGTCAAA 660
Db 601 ATCAGCTAGGAGTTGCTGCTGAGAGGCTTCTTATCTGCTGAGGAAATCTGTCAAA 660
QY 661 TTCAAGAACTATCGCCGACAAAATAGGATACACATTCGAAGAACAACTGGGTACCTTC 720
Db 661 TTCAAGAACTATCGCCGACAAAATAGGATACACATTCGAAGAACAACTGGGTACCTTC 720
QY 721 TGTAAAGCAATCCAGGAATCAAAATATACACAAAGCAACACAGCAACACTAACAGTCA 780
Db 721 TGTAAAGCAATCCAGGAATCAAAATATACACAAAGCAACACAGCAACACTAACAGTCA 780
QY 781 AGCAAGTCAAGTAATGACATGATGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 840
Db 781 AGCAAGTCAAGTAATGACATGATGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 840
QY 841 TCACACATGATGAATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCCGAAC 900
Db 841 TCACACATGATGAATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCCGAAC 900
QY 901 AGCTAGAGGTTGAGTGCCACACGGAGATCATACATTCATCCCTTACTCTCAAT 960
Db 901 AGCTAGAGGTTGAGTGCCACACGGAGATCATACATTCATCCCTTACTCTCAAT 960
QY 961 GTCTGAAATCGAAGCAAGTCGCTGATTTATCCCTTCTGTTATCGTTCAAAACCATTG 1020
Db 961 GTCTGAAATCGAAGCAAGTCGCTGATTTATCCCTTCTGTTATCGTTCAAAACCATTG 1020
QY 1021 GGTACACAGATTCAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
Db 1021 GGTACACAGATTCAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
QY 1081 CCGCAACCTGCACAAAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGCTGGT 1140
Db 1081 CCGCAACCTGCACAAAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGCTGGT 1140
QY 1141 ACAGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTGTC 1200
Db 1141 ACAGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTGTC 1200
QY 1201 GAAAGATTCATCTGAACTGTTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1260

1201 GAAAGATTCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATCAAAAACAGAGAG 1260
1261 TGTTCACACACTTAACTGCTAAAAAGAAAATGTTGCTCTCGTGACCAAGAAATTTTA 1320
1261 TGTTCACACACTTAACTGCTAAAAAGAAAATGTTGCTCTCGTGACCAAGAAATTTTA 1320
1321 TGATAAAGCATATAATCTGTTAACTCATAAAGCCTTGTGTTGNAATAAAGGCTG 1380
1321 TGATAAAGCATATAATCTGTTAACTCATAAAGCCTTGTGTTGNAATAAAGGCTG 1380
1381 TAATTCGTGATTTCCAAAGCCTTAGACAAATATTAAGAAACGTTGAATGATCGACTAA 1440
1381 TAATTCGTGATTTCCAAAGCCTTAGACAAATATTAAGAAACGTTGAATGATCGACTAA 1440
1441 TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCATCCAGAGCG 1500
1441 TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCATCCAGAGCG 1500
1501 ACTTGGCAACCAAAATCTCAAAATTCAGTATATCTGAAGACGAAGTTCGTATTCCTCAAT 1560
1501 ACTTGGCAACCAAAATCTCAAAATTCAGTATATCTGAAGACGAAGTTCGTATTCCTCAAT 1560
1561 AGCTGATAGTATACAAAGTCAGATGTTTACATTTTGTGATGAACATGATATATCACTGA 1620
1561 AGCTGATAGTATACAAAGTCAGATGTTTACATTTTGTGATGAACATGATATATCACTGA 1620
1621 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAGAGATAG 1680
1621 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAGAGATAG 1680
1681 CCTTCTGTATAAGGAAAAAGTTGCGACTCAAGCCTATATAAAGAAAAAGGTATCTTACC 1740
1681 CCTTCTGTATAAGGAAAAAGTTGCGACTCAAGCCTATATAAAGAAAAAGGTATCTTACC 1740
1741 TCCATCTCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACGACTATTTTA 1800
1741 TCCATCTCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACGACTATTTTA 1800
1801 CAATCTGTGAAAGGGGAAAAAGCAATTCACCTGTTTCGACTTCCATATATGTTGAGCA 1860
1801 CAATCTGTGAAAGGGGAAAAAGCAATTCACCTGTTTCGACTTCCATATATGTTGAGCA 1860
1861 TACAGTTAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCAITTAACATAAT 1920
1861 TACAGTTAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCAITTAACATAAT 1920
1921 TAAATTTGCTGTTTGTATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1921 TAAATTTGCTGTTTGTATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1981 TTTGTTTGGCAGCGATTAAGTACTACGTAGAACAACCTGACGACGTCACATTTCTAATGA 2040
1981 TTTGTTTGGCAGCGATTAAGTACTACGTAGAACAACCTGACGACGTCACATTTCTAATGA 2040
2041 TGGATGGGCAATGCCAGTGAGCATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
2041 TGGATGGGCAATGCCAGTGAGCATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
2101 TAAGAACTTCAAGCGGATGAAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
2101 TAAGAACTTCAAGCGGATGAAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTGCGAA 2220
2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTGCGAA 2220
2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACCAATAA 2280
2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACCAATAA 2280
2281 TTTCACTCTTCAAAATTTAGATTAATAGTATCATGGCAGAAAGCAGAAAAATTTACTTGC 2340
2281 TTTCACTCTTCAAAATTTAGATTAATAGTATCATGGCAGAAAGCAGAAAAATTTACTTGC 2340

QY	1261	TGTTTCACACACTTAACTGCTTAAAGGAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1320	Db	2341	GTGTTAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAATAAAC	2389
Db	1261	TGTTTCACACACTTAACTGCTTAAAGGAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1320				
QY	1321	TGATAAAGCATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTTCGNAAATAAGGCTCG	1380				
Db	1321	TGATAAAGCATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTTCGNAAATAAGGCTCG	1380				
QY	1381	TAATTCGTATTTCAAGCCTTAGACAAATATTAGAAACGCTTGAATGATCGACTAA	1440				
Db	1381	TAATTCGTATTTCAAGCCTTAGACAAATATTAGAAACGCTTGAATGATCGACTAA	1440				
QY	1441	TAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAACCATCCAGAGCG	1500				
Db	1441	TAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAACCATCCAGAGCG	1500				
QY	1501	ACTTGGCAAAACCAATTTCTCAAAATTTAGTATCTAGAGACGAAGTTGCTATTGCTCAATT	1560				
Db	1501	ACTTGGCAAAACCAATTTCTCAAAATTTAGTATCTAGAGACGAAGTTGCTATTGCTCAATT	1560				
QY	1561	AGCTGATAAGTATACAAACGTCAGATGTTACATTTTGTATGAACATGATATATCACTGA	1620				
Db	1561	AGCTGATAAGTATACAAACGTCAGATGTTACATTTTGTATGAACATGATATATCACTGA	1620				
QY	1621	TGAAGGATGCTATATGTAAGCCTCATATGGGCCATAGTCACCTGATTTGGAAGATAG	1680				
Db	1621	TGAAGGATGCTATATGTAAGCCTCATATGGGCCATAGTCACCTGATTTGGAAGATAG	1680				
QY	1681	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATATAAGAAAAAGATATCTTACC	1740				
Db	1681	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATATAAGAAAAAGATATCTTACC	1740				
QY	1741	TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA	1800				
Db	1741	TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA	1800				
QY	1801	CAATCGTGTGAAGGGAAGAAACGAATTCCTGCTGCTCCATATATGTTGAGCA	1860				
Db	1801	CAATCGTGTGAAGGGAAGAAACGAATTCCTGCTGCTCCATATATGTTGAGCA	1860				
QY	1861	TACAGTTGAGGTTTAAAGCGGTAAATTTGATTAATTCCTCATAGGATCATACCATATAT	1920				
Db	1861	TACAGTTGAGGTTTAAAGCGGTAAATTTGATTAATTCCTCATAGGATCATACCATATAT	1920				
QY	1921	TAAATTTGCTGTTTGTATGATCACAATCAAGCTCAATGGCTATATCTTGGAGA	1980				
Db	1921	TAAATTTGCTGTTTGTATGATCACAATCAAGCTCAATGGCTATATCTTGGAGA	1980				
QY	1981	TTTGTTCGACGATTAAGTACTAGCTAGAACACCTGACCAACCTCCACATTTCTAATGA	2040				
Db	1981	TTTGTTCGACGATTAAGTACTAGCTAGAACACCTGACCAACCTCCACATTTCTAATGA	2040				
QY	2041	TGGATGGGGCAATGCACTGAGCATGTTGTAGGCAAGAAAGACCAACGATGAAGATCCAAA	2100				
Db	2041	TGGATGGGGCAATGCACTGAGCATGTTGTAGGCAAGAAAGACCAACGATGAAGATCCAAA	2100				
QY	2101	TAAGAACTTCAAGGGATGAGAGCCAGTAGAGCAACCTGCTGAGCAGAGTCCC	2160				
Db	2101	TAAGAACTTCAAGGGATGAGAGCCAGTAGAGCAACCTGCTGAGCAGAGTCCC	2160				
QY	2161	TCAAGTAGAGCTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGAA	2220				
Db	2161	TCAAGTAGAGCTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGAA	2220				
QY	2221	AGTAAAGGATTTAGTGAAGGCAATGCAACAGAACTTAGCTGTTTACGAATAA	2280				
Db	2221	AGTAAAGGATTTAGTGAAGGCAATGCAACAGAACTTAGCTGTTTACGAATAA	2280				
QY	2281	TTTGTACTTCTCAAAATTTATGATAAATAATGTTATGCGAGAGCAAGAAAAATTAATTGC	2340				
Db	2281	TTTGTACTTCTCAAAATTTATGATAAATAATGTTATGCGAGAGCAAGAAAAATTAATTGC	2340				
QY	2341	GTGTTAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAATAAAC	2389				

Db	2341	GTGTTAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAATAAAC	2389
RESULT 4			
ABX06886			
ID	ABX06886	standard; DNA; 2406 BP.	
XX	AC	ABX06886;	
XX	AC		
DT	27-OCT-2003	(revised)	
DT	11-FEB-2003	(first entry)	
XX	S. pneumoniae	type 4 strain coding region #1174.	
DE			
XX	Gene; ds;	bacterial meningitis; pneumonia; sepsis; otitis media;	
KW	ear infection;	antifungal; antibacterial; immunostimulant;	
KW	auditory;	respiratory; gene therapy; vaccine.	
XX	Streptococcus	pneumoniae; type 4 strain.	
OS			
XX	W0200277021-A2.		
PN			
XX	03-OCT-2002.		
PD			
XX	27-MAR-2002;	2002WO-IB002163.	
XX	27-MAR-2001;	2001GB-00007658.	
PF			
XX	(CHIR-) CHIRON SPA.		
XX	(GENO-) INST GENOMIC RES.		
PA			
XX	Masighani V, Tettelin H, Fraser C;		
PI			
XX	WPI; 2003-040579/03.		
DR	P-PSDB; ABU01598.		
DR			
XX	New proteins and nucleic acid molecules from Streptococcus pneumoniae,		
XX	useful as medicaments for treating or preventing a disease or infection		
PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or		
PT	ear infection.		
XX			
FS	Claim 6; SEQ ID NO 2347; 56pp; English.		
XX			
CC	The invention relates to a protein comprising or having at least 50%		
CC	identity to any of the 2469 amino acid sequences, identified in the		
CC	specification (available on a computer readable format), or its fragment,		
CC	expressed from 2469 of 2489 identified DNA coding regions from the		
CC	Streptococcus pneumoniae type 4 strain genomic sequence appearing as		
CC	AB56454. Also included are an antibody which binds one of the proteins,		
CC	treating a patient by administering the protein, DNA or antibody (in a		
CC	composition), a kit comprising first and second primers, which are the		
CC	nucleic acid cited above or fragments between nucleotides 8-100 of a		
CC	sequence not defined in the specification, for amplifying a target		
CC	sequence contained within a Streptococcus nucleic acid sequence, where		
CC	the first primer is substantially complementary to the target sequence		
CC	and the second primer is substantially complementary to the complement of		
CC	the target sequence, and where the parts of the primers having		
CC	substantial complementarity define the termini of the target sequence to		
CC	be amplified, assay comprising contacting a test compound with the		
CC	protein, and determining whether the test compound binds to the protein		
CC	and a Streptococcus pneumoniae bacterium, where one or more genes		
CC	encoding the proteins has been rendered inactive. The proteins, nucleic		
CC	acid molecules, antibody and compositions are useful as medicaments for		
CC	treating or preventing a disease or infection due to streptococcus		
CC	bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis		
CC	media or ear infection. They are also useful in developing vaccines,		
CC	diagnostics and antibiotics. The methods are useful for identifying		
CC	immunodominant proteins. The present sequence is one of the 2489		
CC	identified coding region from the genomic sequence. Note: The sequence		
CC	data for this patent did not form part of the printed specification, but		
CC	was obtained in electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to		

[illegible]

Db 2058 TGGATGGGCAATGCCAGTCAGCATGTGTTAGSCAAGAAAGACCACAGTGAAGATCCAAA 2117
 QY 2101 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
 Db 2118 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAGTCCC 2177
 QY 2161 TCAAGTAGAGACTGAAAGAGTGAAGCCCAACTCAAGAGAGAGAGAGTGTGCTTGGGAA 2220
 Db 2178 TCAAGTAGAGACTGAAAGAGTGAAGCCCAACTCAAGAGAGAGAGAGTGTGCTTGGGAA 2237
 QY 2221 AGTAACGGATCTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 Db 2238 AGTAACGGATCTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2297
 QY 2281 TTTGACTCTTCAATATGATGAATCAATAGTATCATGCGAGAGACAGAAAATTTACTTGC 2340
 Db 2298 TTTGACTCTTCAATATGATGAATCAATAGTATCATGCGAGAGACAGAAAATTTACTTGC 2357
 QY 2341 GTTGTAAAAGGAAGTAACTCTTCACTGTAAGTAAAGGAAAAATAAAC 2389
 Db 2358 GTTGTAAAAGGAAGTAACTCTTCACTGTAAGTAAAGGAAAAATAAAC 2406

RESULT 5

AAA47604
 ID AAA47604 standard; DNA; 2451 BP.
 XX AC
 XX AAA47604;
 XX 20-OCT-2000 (first entry)
 XX DB
 XX Recombinant variant of Sp36 gene (Sp36A) of *S. pneumoniae*.
 XX DB
 XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX DB
 XX Streptococcus pneumoniae.

Key Location/Qualifiers
 FH 1. 2451
 FT CDS
 FT /*tag= a
 FT /product= "Sp36A polypeptide"

WO2000037105-A2.

29-JUN-2000.

21-DEC-1999; 99WO-US030390.

21-DEC-1998; 98US-0113048P.

(MEDI-) MEDIMUNE INC.

Johnson LS, Koenig S, Adamou JE;

WPI; 2000-452129/39.

P-P5DB; AAB01468.

Vaccine useful for prophylaxis and treatment of pneumococcal infections such as otitis media, nasopharyngeal and bronchial infections, comprises Streptococcus pneumoniae proteins.

Disclosure; Page 64-65; 70pp; English.

XX Although a number of proteins have been suggested as being involved in the pathogenicity of Streptococcus pneumoniae, there still remains a need to identify polypeptides having epitopes in common from various strains of *S. pneumoniae* in order to utilize such polypeptides in vaccines to protect against a wide variety of *S. pneumoniae*. New vaccine compositions are described which comprise a Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino acids in length that comprise at least one

CC histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody directed against these features. The vaccine is useful in protecting CC against infection by Streptococcus pneumoniae. The vaccine composition comprising antibodies to is useful for passive immunization for treating CC pneumococcal infections which includes otitis media, nasopharyngeal and CC bronchial infections
 XX
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 0 U; 1 Other;
 Query Match 100.0%; Score 2388; DB 3; Length 2451;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGGAGCTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTGTTCCTA 60
 Db 60 TTCTTACGAGTTGGGAGCTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTGTTCCTA 119
 QY 61 TATAGATGAAAAACAGCGACGCAAAAAACGGAGATTTCGACTCCTGATGAGGTACAA 120
 Db 120 TATAGATGAAAAACAGCGACGCAAAAAACGGAGATTTCGACTCCTGATGAGGTACAA 179
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGTCAC 180
 Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGTCAC 239
 QY 181 TTCATGCGGACCACTATCATTTATTAATGTTAGGTTCCTTATGACGCTATCATCAG 240
 Db 240 TTCATGCGGACCACTATCATTTATTAATGTTAGGTTCCTTATGACGCTATCATCAG 299
 QY 241 TGAAGAATTACTCATGAAGATCCAACTAATGAAGTAAAGATGAGATATGTTAATGA 300
 Db 300 TGAAGAATTACTCATGAAGATCCAACTAATGAAGTAAAGATGAGATATGTTAATGA 359
 QY 301 GGTCAAGGTTGGATATGTTATCAAGTAGTAGTAAAAATACCTATTTACCTAAAGGATGC 360
 Db 360 GGTCAAGGTTGGATATGTTATCAAGTAGTAGTAAAAATACCTATTTACCTAAAGGATGC 419
 QY 361 TGCCACGCGGATAAACGTCGCTACAAAGAGGAATCAATCGACAAAAACAGAGCATAG 420
 Db 420 TGCCACGCGGATAAACGTCGCTACAAAGAGGAATCAATCGACAAAAACAGAGCATAG 479
 QY 421 TCACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTCTGCTTGGCACCCTTCCGA 480
 Db 480 TCACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTCTGCTTGGCACCCTTCCGA 539
 QY 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 Db 540 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 599
 QY 541 TGGTATGCTTATATCGTTCCTCATGGAGATCATACCATTTACATTCCTAAGNATGAGTT 600
 Db 600 TGGTATGCTTATATCGTTCCTCATGGAGATCATACCATTTACATTCCTAAGNATGAGTT 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCTCTATCTGCTGAGGAAATCTGTCAA 660
 Db 660 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCTCTATCTGCTGAGGAAATCTGTCAA 719
 QY 661 TTCAGAAACCTATCGCCGACAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC 720
 Db 720 TTCAGAAACCTATCGCCGACAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC 779
 QY 721 TGTAAAGCAATCCAGGAATCAAAATACCTAAACAGCAACACAGCACTAACAGTCA 780
 Db 780 TGTAAAGCAATCCAGGAATCAAAATACCTAAACAGCAACACAGCACTAACAGTCA 839
 QY 781 AGCAAGTCAAAGTAAATGACATTTAGTCTCTTGAACAGCTCTTCAAACTCCCTTTGAG 840
 Db 840 AGCAAGTCAAAGTAAATGACATTTAGTCTCTTGAACAGCTCTTCAAACTCCCTTTGAG 899
 QY 841 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGAGTCGAC 900
 Db 900 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGAGTCGAC 959

901 AGCTAGAGGTGTTCCAGTGGCCACACGAGATCAITACCACTTCATCCCTTACTCTCAAT 960
960 AGCTAGAGGTGTTCCAGTGGCCACACGAGATCAITACCACTTCATCCCTTACTCTCAAT 1019
961 GTCTGAATTTGGAAGAAAGAAATCGCTCGTATATATTCCTCCTTCTGTTATCGTTTCAAAACCATTTG 1020
1020 GTCTGAATTTGGAAGAAAGAAATCGCTCGTATATATTCCTCCTTCTGTTATCGTTTCAAAACCATTTG 1079
1021 GGTACACAGATTCAAGGCGAGAAACCAAGTCCACAACCGACTCCGGNACCTAGTCCAGG 1080
1080 GGTACACAGATTCAAGGCGAGAAACCAAGTCCACAACCGACTCCGGNACCTAGTCCAGG 1139
1081 CCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTCAAGTGGT 1140
1140 CCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTCAAGTGGT 1199
1141 ACCGAAAGTTGGGGAAGGATATGTTTTCGAAAGAAAGGGGATCTCTCGTTATGCTTTGC 1200
1200 ACCGAAAGTTGGGGAAGGATATGTTTTCGAAAGAAAGGGGATCTCTCGTTATGCTTTGC 1259
1201 GAAAGATTATACCATCTGAAACTGTTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG 1260
1260 GAAAGATTATACCATCTGAAACTGTTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG 1319
1261 TGTTCACACATTTAACTCTGAAAGCAAAATGTTGCTCCTCGTGACCAAGATTTTA 1320
1320 TGTTCACACATTTAACTCTGAAAGCAAAATGTTGCTCCTCGTGACCAAGATTTTA 1379
1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGCTG 1380
1380 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGCTG 1439
1381 TAATCTGATTTCCAAAGCCTTAGACAAATATTTAGAACGCTTCAATGATCAATCGACTAA 1440
1440 TAATCTGATTTCCAAAGCCTTAGACAAATATTTAGAACGCTTCAATGATCAATCGACTAA 1499
1441 TAAAGAAAATTTGGTAGATGATTTATGGCANTCTAGCAACCAATACCCATCCAGAGCG 1500
1500 TAAAGAAAATTTGGTAGATGATTTATGGCANTCTAGCAACCAATACCCATCCAGAGCG 1559
1501 ACTTGGCAACCAAAATCTCAATTCAGTATATCTGAAGAGAAAGTTCGTTATGCTCAAT 1560
1560 ACTTGGCAACCAAAATCTCAATTCAGTATATCTGAAGAGAAAGTTCGTTATGCTCAAT 1619
1561 AGCTGATAAGTATACAACTGATGGTTACATTTTGTGATGAACATGATATATCATGCA 1620
1620 AGCTGATAAGTATACAACTGATGGTTACATTTTGTGATGAACATGATATATCATGCA 1679
1621 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAGAGATAG 1680
1680 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAGAGATAG 1739
1681 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCCTATACCTGATTAAGAAAAGGTTATCTACC 1740
1740 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCCTATACCTGATTAAGAAAAGGTTATCTACC 1799
1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGCAGCTATTTA 1800
1800 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGCAGCTATTTA 1859
1801 CAATCGTGAAGGAAAAGTTCGAATTCATCGTTTCGATTCCTCATATATGTTGAGCA 1860
1860 CAATCGTGAAGGAAAAGTTCGAATTCATCGTTTCGATTCCTCATATATGTTGAGCA 1919
1861 TACAGTTGAGGTTAAACCGGTAATTTGATTTTCTCATAGGATCATACCAATATAT 1920
1920 TACAGTTGAGGTTAAACCGGTAATTTGATTTTCTCATAGGATCATACCAATATAT 1979
1921 TAAATTTGCTGTTTGTATGATCAACATACCAAGCTCCAAATGCTATACCTTCGAGA 1980
1980 TAAATTTGCTGTTTGTATGATCAACATACCAAGCTCCAAATGCTATACCTTCGAGA 2039
1981 TTTGTTGGCAGGATTAAGTACTAGTAGAACACCCCTGACGAACGTTCCACATTTCTAATGA 2040

Db 2040 TTTGTTTGGACGATTAAGTACTAGTAGAACACCCCTGACGAAGTCCACATTTCTAATGA 2099
Qy 2041 TGGATGGGCAATGCCAGTGAAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA 2100
Db 2100 TGGATGGGCAATGCCAGTGAAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA 2159
Qy 2101 TAAAGACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTCTGAGCCAGAAGTCCC 2160
Db 2160 TAAAGACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTCTGAGCCAGAAGTCCC 2219
Qy 2161 TCAAGTAGAGACTCAAAAGCTAGAGCCCAACTCAAGAGACGAGAGTTTGTCTGGAA 2220
Db 2220 TCAAGTAGAGACTCAAAAGCTAGAGCCCAACTCAAGAGAGCAGAGTTTGTCTGGAA 2279
Qy 2221 AGTAACGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db 2280 AGTAACGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2339
Qy 2281 TTTGACTCTTCAAATATATGATTAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340
Db 2340 TTTGACTCTTCAAATATATGATTAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2399
Qy 2341 GTTGTAAAAGGAGTAATCTCTCATCTGTAAGTAAAGAAAAATAAAC 2389
Db 2400 GTTGTAAAAGGAGTAATCTCTCATCTGTAAGTAAAGAAAAATAAAC 2448

RESULT 6

AAV52227

ID AAV52227 standard; DNA; 8195 BP.

XX AAV52227;

XX 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:94.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

XX computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019588.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillion PJ, Rosen CA, Barash SC, Fannon M;

XX Dougherty BA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

XX polynucleotide sequences - useful in diagnostic kits and assays, and

XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 727-732; 1409pp; English.

XX The present invention describes a computer readable medium which has the

XX nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded

XX on it, or a representative fragment or a sequence at least 95% identical

XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

XX to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus

XX pneumoniae. The present invention also describes an isolated nucleic acid

XX molecule encoding a homologue of any of the fragments of the S.pneumoniae

XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

XX by a process comprising: (a) screening a genomic DNA library using as a

CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1 to	
CC	391, identifying members of the library which contain sequences that	
CC	hybridize to the target sequence and isolating the nucleic acid molecules	
CC	from the members; or (b) isolating mRNA, DNA or cDNA produced from an	
CC	organism, amplifying nucleic acid molecules whose nucleotide sequence is	
CC	homologous to amplification primers derived from the fragment of the S.	
CC	pneumoniae genome to prime the amplification and isolating the amplified	
CC	sequences. The computer readable medium can be used in a computer-based	
CC	system for identifying fragments of the S. pneumoniae genome of	
CC	commercial importance, or expression modulating fragments of the S.	
CC	pneumoniae genome. Products from the present invention can be used in	
CC	diagnosis kits and assays, and pharmaceutical compositions and vaccines	
CC	for S. pneumoniae	
xx		
SQ	Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;	
	Query Match 100.0%; Score 2388; DB 2; Length 8195;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 TTCTTAGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATATCTGTTCTCTA 60	
Db	3053 TTCTTAGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATATCTGTTCTCTA 3112	
Qy	61 TATAGATGGAACCAAGCGACGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 120	
Db	3113 TATAGATGGAACCAAGCGACGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 3172	
Qy	121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGTCAC 180	
Db	3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGTCAC 3232	
Qy	181 TTCACATGGGACACATATATTATCAATGTTAGGTTCTTATGAGCTGATCATCAG 240	
Db	3233 TTCACATGGGACACATATATTATCAATGTTAGGTTCTTATGAGCTGATCATCAG 3292	
Qy	241 TGAAGATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAATGA 300	
Db	3293 TGAAGATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAATGA 3352	
Qy	301 GGTCAAGGGTGGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360	
Db	3353 GGTCAAGGGTGGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 3412	
Qy	361 TGCCACCGGATACGTCCTGACAAAGAGAGAAATCAATCCACAAACCAAGACATAG 420	
Db	3413 TGCCACCGGATACGTCCTGACAAAGAGAGAAATCAATCCACAAACCAAGACATAG 3472	
Qy	421 TCAACATCTGAGGTGGAACTCCAAGAAACGATGGTGTGTCCTTGGCACGTTCCGA 480	
Db	3473 TCAACATCTGAGGTGGAACTCCAAGAAACGATGGTGTGTCCTTGGCACGTTCCGA 3532	
Qy	481 AGGACGCTATACACAGATGATGGTTATATCTTTAATGTTCTGATATCATGAGGATAC 540	
Db	3533 AGGACGCTATACACAGATGATGGTTATATCTTTAATGTTCTGATATCATGAGGATAC 3592	
Qy	541 TGGTATGCTTATATCGTTCTCTCATGAGATCATATACCATTAATCTCTAAGAAAGAGTT 600	
Db	3593 TGGTATGCTTATATCGTTCTCTCATGAGATCATATACCATTAATCTCTAAGAAAGAGTT 3652	
Qy	601 ATCAGCTAGCGAGTTGGCTGCTGACAAAGCTTCTCTATCTGGTTCGAGGAAATCTGTCAA 660	
Db	3653 ATCAGCTAGCGAGTTGGCTGCTGACAAAGCTTCTCTATCTGGTTCGAGGAAATCTGTCAA 3712	
Qy	661 TTCAAGAACTTATCGCCGCAAAATAGCGATATACCTTCAAGAACCAACTGGGTACCTTC 720	
Db	3713 TTCAAGAACTTATCGCCGCAAAATAGCGATATACCTTCAAGAACCAACTGGGTACCTTC 3772	
Qy	721 TGTAGCAATCCAGGAACCTACAATACTACCAAGCAACCAAGCAACCACTAACAGTCA 780	
Db	3773 TGTAGCAATCCAGGAACCTACAATACTACCAAGCAACCAAGCAACCACTAACAGTCA 3832	
Qy	781 AGCAAGTCAAGTAATGACATTGATAGTCTCTTTGAAACAGCTCTACAACCTGCTTTGAG 840	

Db	3833 AGCAAGTCAAGTAATGACATTGATAGTCTCTTTGAAACAGCTCTACAACTGCTTTGAG 3892	
Qy	841 TCAACGACATGTAGATCTGATCGCTTGTCTTTGATCCAGCAAAATCAAACTGCAAC 900	
Db	3893 TCAACGACATGTAGATCTGATCGCTTGTCTTTGATCCAGCAAAATCAAACTGCAAC 3952	
Qy	901 AGCTAGAGTGTGTCAGTGCCACACGAGATCAATACCACTCATCCCTTACTCTCAAT 960	
Db	3953 AGCTAGAGTGTGTCAGTGCCACACGAGATCAATACCACTCATCCCTTACTCTCAAT 4012	
Qy	961 GTCTGAATTCGGAAGAACGAATCGCTCGTATATTTCCCTTGGTATCGTTATCGTTCAACCATG 1020	
Db	4013 GTCTGAATTCGGAAGAACGAATCGCTCGTATATTTCCCTTGGTATCGTTATCGTTCAACCATG 4072	
Qy	1021 GGTACACAGATTCAAGGCCAGAACCAACCAAGTCCCAACCCGACTCCGGAACCTAGTCAGG 1080	
Db	4073 GGTACACAGATTCAAGGCCAGAACCAACCAAGTCCCAACCCGACTCCGGAACCTAGTCAGG 4132	
Qy	1081 CCCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTTCTTTGGTGTAGTCACTGGT 1140	
Db	4133 CCCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTTCTTTGGTGTAGTCACTGGT 4192	
Qy	1141 ACGAAAAGTTGGGGAAGGATATGTTTCGAGAAAGGCGATCTCTCGTTATGTTCTTTCG 1200	
Db	4193 ACGAAAAGTTGGGGAAGGATATGTTTCGAGAAAGGCGATCTCTCGTTATGTTCTTTCG 4252	
Qy	1201 GAAAGATTACCTCTGAAACTGTTTAAATCTTTGAAAGCAAGTATCAAAACCAAGAGAG 1260	
Db	4253 GAAAGATTACCTCTGAAACTGTTTAAATCTTTGAAAGCAAGTATCAAAACCAAGAGAG 4312	
Qy	1261 TGTTTTCACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTCGTGACCAAGAAATTTTA 1320	
Db	4313 TGTTTTCACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTCGTGACCAAGAAATTTTA 4372	
Qy	1321 TGATAAGCATATTAATCTGTTTAACTGAGGCTCATAAAGCCTTGTGTAATTAAGGGTCG 1380	
Db	4373 TGATAAGCATATTAATCTGTTTAACTGAGGCTCATAAAGCCTTGTGTAATTAAGGGTCG 4432	
Qy	1381 TAAATCTGATTTCCAAAGCCTTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440	
Db	4433 TAAATCTGATTTCCAAAGCCTTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 4492	
Qy	1441 TAAAGAAAATTTGGTGTAGATGATTTATTTGGCATTTCTAGACCAATATCCCATCCAGAGG 1500	
Db	4493 TAAAGAAAATTTGGTGTAGATGATTTATTTGGCATTTCTAGACCAATATCCCATCCAGAGG 4552	
Qy	1501 ACTTGGCAAAACCAATCTCTCAAAATTTGATGATGATGATGATGATGATGATGATGATG 1560	
Db	4553 ACTTGGCAAAACCAATCTCTCAAAATTTGATGATGATGATGATGATGATGATGATGATG 4612	
Qy	1561 AGCTGATAAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATTAATCAGTGA 1620	
Db	4613 AGCTGATAAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATTAATCAGTGA 4672	
Qy	1621 TGAAGGAGATGATATGATTAACCCCTCATATGCGCCATAGTCTACTGATTTGGAAGATAG 1680	
Db	4673 TGAAGGAGATGATATGATTAACCCCTCATATGCGCCATAGTCTACTGATTTGGAAGATAG 4732	
Qy	1681 CCTTTCTGATAAGGAAAAGTTGAGCTCAAGCCTTATCTAAAGAAAAGAGTATCTCTACC 1740	
Db	4733 CCTTTCTGATAAGGAAAAGTTGAGCTCAAGCCTTATCTAAAGAAAAGAGTATCTCTACC 4792	
Qy	1741 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCGACTATTTTA 1800	
Db	4793 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCGACTATTTTA 4852	
Qy	1801 CAATCGTGTGAAAGGGGAAAACGAAATCTCACTCTGCTTCCATATATGTTGTGAGCA 1860	
Db	4853 CAATCGTGTGAAAGGGGAAAACGAAATCTCACTCTGCTTCCATATATGTTGTGAGCA 4912	
Qy	1861 TACAGTTGAGGTTTAAAGACGGTAAATTTGATTTTCTCTATAAGGATCATTTACATATAT 1920	

Db 4913 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCATTAACCAATAT 4972
 Qy 1921 TAAATTTGCTTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
 Db 4973 TAAATTTGCTTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 5032
 Qy 1981 TTTGTTTGGCAGATTAAAGTACTAGTATAGAACACCTCGACGAACGCTCCACATTTCTAATGA 2040
 Db 5033 TTTGTTTGGCAGATTAAAGTACTAGTATAGAACACCTCGACGAACGCTCCACATTTCTAATGA 5092
 Qy 2041 TGGATGGGCAATCCAGTGGAGCTGTGTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
 Db 5093 TGGATGGGCAATCCAGTGGAGCTGTGTAGGCAAGAAAGACACAGTGAAGATCCAAA 5152
 Qy 2101 TAAAGCTTTCAAAGCGGATGAAGAGCCAGTAGAGAAACACCTCGAGCCAGAGTCCC 2160
 Db 5153 TAAAGCTTTCAAAGCGGATGAAGAGCCAGTAGAGAAACACCTCGAGCCAGAGTCCC 5212
 Qy 2161 TCAAGTAGAGACTGAAAAGGTAGAGAGCCCAACTCAAAGAGAGCAAGATTTTGTTCGGA 2220
 Db 5213 TCAAGTAGAGACTGAAAAGGTAGAGAGCCCAACTCAAAGAGAGCAAGATTTTGTTCGGA 5272
 Qy 2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAAGAAACTCTAGCTGGTTTACGAATAA 2280
 Db 5273 AGTAACGGATTCTAGTCTGAAAGCCAATGCAAGAAACTCTAGCTGGTTTACGAATAA 5332
 Qy 2281 TTTGACTCTTCAAATATTGGATAACAATAGTATCATGGCAGAGCAAGAAATAAAC 2389
 Db 5333 TTTGACTCTTCAAATATTGGATAACAATAGTATCATGGCAGAGCAAGAAATAAAC 5392
 Qy 2341 GTTGTAAAAGGAAGTAATCCTTCATCTGTAAGTAAAGTAAGGAAAAATAAAC 2389
 Db 5393 GTTGTAAAAGGAAGTAATCCTTCATCTGTAAGTAAAGTAAGGAAAAATAAAC 5441

RESULT 7

ABSS56454_09/c
 Continuation (10 of 22) of ABSS56454 from base 900001 (Streptococcus pneumoniae type 4 sd
 WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

Fragment Name Begin End

ABSS56454_01 1 110000

ABSS56454_02 100001 210000

ABSS56454_03 200001 310000

ABSS56454_04 300001 410000

ABSS56454_05 400001 510000

ABSS56454_06 500001 610000

ABSS56454_07 600001 710000

ABSS56454_08 700001 810000

ABSS56454_09 800001 910000

ABSS56454_10 900001 1010000

ABSS56454_11 1000001 1110000

ABSS56454_12 1200001 1310000

ABSS56454_13 1300001 1410000

ABSS56454_14 1400001 1510000

ABSS56454_15 1500001 1610000

ABSS56454_16 1600001 1710000

ABSS56454_17 1700001 1810000

ABSS56454_18 1800001 1910000

ABSS56454_19 1900001 2010000

ABSS56454_20 2000001 2110000

ABSS56454_21 2100001 2162598

Query Match

Best Local Similarity 100.0%; Score 2388; DB 7; Length 110000;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTTCTACGATTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCTTA 60
 Db 106950 TTTCTACGATTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCTTA 106891
 Qy 61 TATAGATGAAACACAGCGACCCAAAAACCGAGAAATTTGATCTCTGATAGGTTAGCAA 120
 Db 105810 ACRAAAAGTTGGGAGAGGATATGTTATTCGAAGAAAGGCGATCTCTCGTTATGCTTTGC 1200
 Qy 1141 ACRAAAAGTTGGGAGAGGATATGTTATTCGAAGAAAGGCGATCTCTCGTTATGCTTTGC 105751

Db 106890 TATAGATGGAATAACAAGCGACGCAAAAAACGAGAAATTTGATCTCTGATAGGTTAGCAA 106831
 Qy 121 GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAAACAAGCAAGGCTATGTCCAC 180
 Db 106830 CGCTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAAACAAGCAAGGCTATGTCCAC 106771
 Qy 181 TTTACATGGCGACCACTATCATTTATTAATGGAAGGTTCTTTATGACGCTATCATCAG 240
 Db 106770 TTTACATGGCGACCACTATCATTTATTAATGGAAGGTTCTTTATGACGCTATCATCAG 106711
 Qy 241 TGAAGAAATTAATCTCAATGAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA 300
 Db 106710 TGAAGAAATTAATCTCAATGAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA 106651
 Qy 301 GGTCAAGGGTGGATGATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
 Db 106650 GGTCAAGGGTGGATGATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 106591
 Qy 361 TGCCCAACGCGATTAACGCTCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
 Db 106590 TGCCCAACGCGATTAACGCTCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 106531
 Qy 421 TCAACATCGTGAAGGTGGAATCTCCAAAGAACGATGCTGCTGCTTGGCCTTGGCAGCTTCGCA 480
 Db 106530 TCAACATCGTGAAGGTGGAATCTCCAAAGAACGATGCTGCTGCTTGGCCTTGGCAGCTTCGCA 106471
 Qy 481 AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 Db 106470 AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 106411
 Qy 541 TGTGATGCTTATATGTTTCTCATGAGATCAATACCATATGATTCCTTAAGAAATGAGTT 600
 Db 106410 TGTGATGCTTATATGTTTCTCATGAGATCAATACCATATGATTCCTTAAGAAATGAGTT 106351
 Qy 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGTCGAGGAATCTGTCAA 660
 Db 106350 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGTCGAGGAATCTGTCAA 106291
 Qy 661 TTAAGAACTTATCGCGCAAAAATAGCGATAACACTTCAAGAAACAACTGGGTGACTTTC 720
 Db 106290 TTAAGAACTTATCGCGCAAAAATAGCGATAACACTTCAAGAAACAACTGGGTGACTTTC 106231
 Qy 721 TGTAAAGCAATCCAGGAATCAAAATACAAAGAGCAACAGCAACCACTAACAGTCA 780
 Db 106230 TGTAAAGCAATCCAGGAATCAAAATACAAAGAGCAACAGCAACCACTAACAGTCA 106171
 Qy 781 AGCAAGTCAAAGTAATGATGATGCTCTTGAAGACAGCTCTACAAAGCTGCTTTGAG 840
 Db 106170 AGCAAGTCAAAGTAATGATGATGCTCTTGAAGACAGCTCTACAAAGCTGCTTTGAG 106111
 Qy 841 TCAACGATGTAGAAATCTGATGGGCTTCTTTGATCCAGCAACAAATCACAAGTCAAC 900
 Db 106110 TCAACGATGTAGAAATCTGATGGGCTTCTTTGATCCAGCAACAAATCACAAGTCAAC 106051
 Qy 901 AGCTAGAGGTGTGAGTCCGACGAGATCAATACCACTTCCCTTCTTCTTCAAT 960
 Db 106050 AGCTAGAGGTGTGAGTCCGACGAGATCAATACCACTTCCCTTCTTCTTCAAT 105991
 Qy 961 GTCTGAATTTGGAAGAACGAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATTTG 1020
 Db 105990 GTCTGAATTTGGAAGAACGAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATTTG 105931
 Qy 1021 GGTACAGATTCAGGCGAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
 Db 105930 GGTACAGATTCAGGCGAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 105871
 Qy 1081 CCCGCAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCACTGCT 1140
 Db 105870 CCCGCAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCACTGCT 105811
 Qy 1141 ACRAAAAGTTGGGAGAGGATATGTTATTCGAAGAAAGGCGATCTCTCGTTATGCTTTGC 1200
 Db 105810 ACRAAAAGTTGGGAGAGGATATGTTATTCGAAGAAAGGCGATCTCTCGTTATGCTTTGC 105751

1201 GAAAGATTACCACTGTAAGAACTGTTAAATACTTTGAAAGCAAGTTATCAAAAACAGAGAG 1260
105750 GAAAGATTACCACTGTAAGAACTGTTAAATACTTTGAAAGCAAGTTATCAAAAACAGAGAG 105691
1261 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
105690 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 105631
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATATAAGCCCTTGTTTGNAAATAAAGGCTCG 1380
105630 TGATAAAGCATATATCTGTTAACTGAGGCTCATATAAGCCCTTGTTTGNAAATAAAGGCTCG 105571
1381 TAATCTGATTTCCAGCCTTAGACAAATTAATTAAGAACGCTTGATGATGATGATGATGATGAT 1440
105570 TAATCTGATTTCCAGCCTTAGACAAATTAATTAAGAACGCTTGATGATGATGATGATGATGAT 105511
1441 TAAAGAAATTTGCTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCCATCCAGAGCG 1500
105510 TAAAGAAATTTGCTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCCATCCAGAGCG 105451
1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGACGAAGTTCGTATGCTCAATTT 1560
105450 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGACGAAGTTCGTATGCTCAATTT 105391
1561 AGCTGATAAGTATACACGTCAGATGTTTATTTGATGAACATGATATTAATCAAGTGA 1620
105390 AGCTGATAAGTATACACGTCAGATGTTTATTTGATGAACATGATATTAATCAAGTGA 105331
1621 TGAAGGATGATATGTAAGCCCTCATATGGCCATAGTCACTGGATGGAAAGATAG 1680
105330 TGAAGGATGATATGTAAGCCCTCATATGGCCATAGTCACTGGATGGAAAGATAG 105271
1681 CCTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTATATCAAGAAAAAGTATCCTTACC 1740
105270 CCTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTATATCAAGAAAAAGTATCCTTACC 105211
1741 TCCATCTCAGCGAGATGTTTAAAGCAATCCAACTGGATAGTGCAGAGCTATTTA 1800
105210 TCCATCTCAGCGAGATGTTTAAAGCAATCCAACTGGATAGTGCAGAGCTATTTA 105151
1801 CAATCGTGTGAAGGGGAAAAACGAATCCACTGCTTGCATTCATATATGTTGAGCA 1860
105150 CAATCGTGTGAAGGGGAAAAACGAATCCACTGCTTGCATTCATATATGTTGAGCA 105091
1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCAATAAGGATCATTAACATAAT 1920
105090 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCAATAAGGATCATTAACATAAT 105031
1921 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
105030 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 104971
1981 TTTGTTGCGAGATTAAGTACTACGTAGAACACCCCTGACGACGATCCCACTTCAATGA 2040
104970 TTTGTTGCGAGATTAAGTACTACGTAGAACACCCCTGACGACGATCCCACTTCAATGA 104911
2041 TGGATGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAGACCAAGTGAAGATCCAAA 2100
104910 TGGATGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAGACCAAGTGAAGATCCAAA 104851
2101 TAAGAACTTCAAGCGGATGAAGCGCAGTGAAGGAAACACCTGCTGAGCCGAAGTCCC 2160
104850 TAAGAACTTCAAGCGGATGAAGCGCAGTGAAGGAAACACCTGCTGAGCCGAAGTCCC 104791
2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGATTTTGTTCGGAA 2220
104790 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGATTTTGTTCGGAA 104731
2221 AGTAAGCGGATTCAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA 2280
104730 AGTAAGCGGATTCAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA 104671

2281 TTTGACTCTTCAAAATATGATTAACAAATAGTATCATGGGAGAGACAGAAAAATTTACTTGC 2340
104670 TTTGACTCTTCAAAATATGATTAACAAATAGTATCATGGGAGAGACAGAAAAATTTACTTGC 104611
2341 GTTGTAAAGGAAGTAATCTTCACTCTGAAGTAAGGAAAAATAAAC 2389
104610 GTTGTAAAGGAAGTAATCTTCACTCTGAAGTAAGGAAAAATAAAC 104562

RESULT 8

ABS56454 10/c

Continuation (11 of 22) of ABS56454 from base 1000001 (Streptococcus pneumoniae type 4 st

WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454

Fragment Name Begin End

WP ABS56454_00 1 110000

WP ABS56454_01 100001 210000

WP ABS56454_02 200001 310000

WP ABS56454_03 300001 410000

WP ABS56454_04 400001 510000

WP ABS56454_05 500001 610000

WP ABS56454_06 600001 710000

WP ABS56454_07 700001 810000

WP ABS56454_08 800001 910000

WP ABS56454_09 900001 1010000

WP ABS56454_10 1000001 1110000

WP ABS56454_11 1100001 1210000

WP ABS56454_12 1200001 1310000

WP ABS56454_13 1300001 1410000

WP ABS56454_14 1400001 1510000

WP ABS56454_15 1500001 1610000

WP ABS56454_16 1600001 1710000

WP ABS56454_17 1700001 1810000

WP ABS56454_18 1800001 1910000

WP ABS56454_19 1900001 2010000

WP ABS56454_20 2000001 2110000

WP ABS56454_21 2100001 2162598

Query Match 100.0%; Score 2388; DB 7; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTTTACAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATTAATCGTGTTCCTTA 60

DB 6950 TTTTACAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATTAATCGTGTTCCTTA 6891

QY 61 TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATAGGTTAGCAA 120

DB 6890 TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATAGGTTAGCAA 6831

QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCTCAAGATAACAGACCAAGGCTATGTCAC 180

DB 6830 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCTCAAGATAACAGACCAAGGCTATGTCAC 6771

QY 181 TTTACATGGCGACCACTATCAATTTACAATGGTAAAGTTTCTTATGACGCTATCATCAG 240

DB 6770 TTTACATGGCGACCACTATCAATTTACAATGGTAAAGTTTCTTATGACGCTATCATCAG 6711

QY 241 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 300

DB 6710 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 6651

QY 301 GGTCAAGGTTGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACTCTTAAGGATGC 360

DB 6650 GGTCAAGGTTGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACTCTTAAGGATGC 6591

QY 361 TGCCCCACGGGATAACGCTCCGTACAAAAGAGAAAAATCAATCGACAAAAACAAGAGCATAG 420

DB 6590 TGCCCCACGGGATAACGCTCCGTACAAAAGAGAAAAATCAATCGACAAAAACAAGAGCATAG 6531

QY 421 TCAACATCGTGAAGGTGGAATCCAAAGAAACGATGCTGCTGTTGCCCTTGCCACGTTCCGA 480

DB 6530 TCAACATCGTGAAGGTGGAATCCAAAGAAACGATGCTGCTGTTGCCCTTGCCACGTTCCGA 6471

481 AGGACGCTATCTACAGATGATGCTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
|||||
6470 AGGACGCTATCTACAGATGATGCTTATATCTTTAATGCTTCTGATATCATAGAGATAC 6411
|||||
541 TGGTATGCTTATATCGTTCTCATGAGATCATATACCAATTAATTCCTAAGAAATGAGTT 600
|||||
6410 TGGTATGCTTATATCGTTCTCATGAGATCATATACCAATTAATTCCTAAGAAATGAGTT 6351
|||||
601 ATCAGTACGAGTGGCTGCTGCTGAGAGCCTTCCTATCTGCTGAGAGAAATCTGTCAA 660
|||||
6350 ATCAGTACGAGTGGCTGCTGCTGAGAGCCTTCCTATCTGCTGAGAGAAATCTGTCAA 6291
|||||
661 TTCAAGAACTTATCGCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
|||||
6290 TTCAAGAACTTATCGCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 6231
|||||
721 TGTAAAGCAATCCAGGAATACAAATCTAACCAAGCAACCAAGCAACCACTAACAGTCA 780
|||||
6230 TGTAAAGCAATCCAGGAATACAAATCTAACCAAGCAACCAAGCAACCACTAACAGTCA 6171
|||||
781 AGCAAGTCAAGTAAATGACATTTGATAGTCTCTTGAAGACAGCTCTACAACTGCTTTGAG 840
|||||
6170 AGCAAGTCAAGTAAATGACATTTGATAGTCTCTTGAAGACAGCTCTACAACTGCTTTGAG 6111
|||||
841 TCAACGACATGTAGAACTCTGATGGCCTTCTCTTGTATCCAGACAAATCAAGTCGAAC 900
|||||
6110 TCAACGACATGTAGAACTCTGATGGCCTTCTCTTGTATCCAGACAAATCAAGTCGAAC 6051
|||||
901 AGCTAGAGTGTGAGTGCCTGCAACGAGATCATTAACCACTTCACTCTCTCAAT 960
|||||
6050 AGCTAGAGTGTGAGTGCCTGCAACGAGATCATTAACCACTTCACTCTCTCAAT 5991
|||||
961 GTCTGAATTTGGAAGACGAATCGCTGATTAATTCCTTCTGCTTATCTGTTCAACATTTG 1020
|||||
5990 GTCTGAATTTGGAAGACGAATCGCTGATTAATTCCTTCTGCTTATCTGTTCAACATTTG 5931
|||||
1021 GGTACCAAGATTCAGGCCAGAAACCAACCAAGTCCACCAACCGACTCGGAACTAGTCCAGG 1080
|||||
5930 GGTACCAAGATTCAGGCCAGAAACCAACCAAGTCCACCAACCGACTCGGAACTAGTCCAGG 5871
|||||
1081 CCGGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGTT 1140
|||||
5870 CCGGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGTT 5811
|||||
1141 AGCAAGTGTGGGAGGATGATGTTTCGAAGAAAGGCATCTCTGTTATGCTTTGCT 1200
|||||
5810 AGCAAGTGTGGGAGGATGATGTTTCGAAGAAAGGCATCTCTGTTATGCTTTGCT 5751
|||||
1201 GAAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260
|||||
5750 GAAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 5691
|||||
1261 TGTTCACACACTTTAATCTGCTTAAAGAAATGTTGCTCTCGTGAACCAAGAAATTTTA 1320
|||||
5690 TGTTCACACACTTTAATCTGCTTAAAGAAATGTTGCTCTCGTGAACCAAGAAATTTTA 5631
|||||
1321 TGATAAGCATATAATCTGTTAACTGAGCTCATTAAGCCTTGTGNAATTAAGGGTCG 1380
|||||
5630 TGATAAGCATATAATCTGTTAACTGAGCTCATTAAGCCTTGTGNAATTAAGGGTCG 5571
|||||
1381 TAAATCTGATTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCACTAA 1440
|||||
5570 TAAATCTGATTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCACTAA 5511
|||||
1441 TAAAGAAAAATTTGGTAGATTTATTTGGCATTTCTAGACAAATTAACCATCCAGAGCG 1500
|||||
5510 TAAAGAAAAATTTGGTAGATTTATTTGGCATTTCTAGACAAATTAACCATCCAGAGCG 5451
|||||
1501 ACTTGGCAACCAATTTCTCAAAATGAGTATGAGTCAAGACGAGTTGCTATGCTCAAT 1560
|||||
5450 ACTTGGCAACCAATTTCTCAAAATGAGTATGAGTCAAGACGAGTTGCTATGCTCAAT 5391
|||||
1561 AGCTGATAAGTATACAAGCTCAGATGGTTTACATTTTGTATGAACATGATAATCAGTGA 1620
|||||

Db 5390 AGCTGATAAGTATACAACGCTCAGATGGTTTACATTTTGTGAAACATGATAATCAGTGA 5331
|||||
QY 1621 TGAAGGAGATGATATGTAAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
|||||
Db 5330 TGAAGGAGATGATATGTAAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 5271
|||||
QY 1681 CTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATCTTAAAGAAAAAGTATCCTTACC 1740
|||||
Db 5270 CTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATCTTAAAGAAAAAGTATCCTTACC 5211
|||||
QY 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
|||||
Db 5210 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 5151
|||||
QY 1801 CAATCTGTGAAAGGGGAAAAACGAATTCACCTCGTTCCACTATATATGTTGTCAGCA 1860
|||||
Db 5150 CAATCTGTGAAAGGGGAAAAACGAATTCACCTCGTTCCACTATATATGTTGTCAGCA 5091
|||||
QY 1861 TACAGTTGAGGTTTAAACCGGTAATTTGATTAATTCCTCAATAAGGATCAATTACCATATAT 1920
|||||
Db 5090 TACAGTTGAGGTTTAAACCGGTAATTTGATTAATTCCTCAATAAGGATCAATTACCATATAT 5031
|||||
QY 1921 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAGA 1980
|||||
Db 5030 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAGA 4971
|||||
QY 1981 TTTGTTTGGCAGATTAAGTACTAGCTAGAACACCTCGACGAACCTCCACATTTCTTAATCA 2040
|||||
Db 4970 TTTGTTTGGCAGATTAAGTACTAGCTAGAACACCTCGACGAACCTCCACATTTCTTAATCA 4911
|||||
QY 2041 TGGATGGGGCAATGCGACGTAGCATGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
|||||
Db 4910 TGGATGGGGCAATGCGACGTAGCATGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 4851
|||||
QY 2101 TAAAGAACTTCAAAGCGGATGAAGCCAGTAGAGCAACACCTGCTGAGCCAGAGTCCC 2160
|||||
Db 4850 TAAAGAACTTCAAAGCGGATGAAGCCAGTAGAGCAACACCTGCTGAGCCAGAGTCCC 4791
|||||
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAAAGCAGAAAGTTTCTCTTCGCA 2220
|||||
Db 4790 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAAAGCAGAAAGTTTCTCTTCGCA 4731
|||||
QY 2221 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAAAATCTAGCTGGTTTACGAATAA 2280
|||||
Db 4730 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAAAATCTAGCTGGTTTACGAATAA 4671
|||||
QY 2281 TTTGACTCTTCAAAATTTATGGATAACAATAGTATCATGGCAGAACGAGAAATTTACTTGC 2340
|||||
Db 4670 TTTGACTCTTCAAAATTTATGGATAACAATAGTATCATGGCAGAACGAGAAATTTACTTGC 4611
|||||
QY 2341 GTTGTAAAGGAAGTAATCTCTTCACTGTAAAGTAAAGAAAAATAAAC 2389
|||||
Db 4610 GTTGTAAAGGAAGTAATCTCTTCACTGTAAAGTAAAGAAAAATAAAC 4562
|||||

RESULT 9

AAA65731

ID AAA65731 standard; DNA; 2523 BP.

XX AAA65731;

XX AC

XX XX

DT 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.

DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal; ds.

XX Streptococcus pneumoniae.

XX OS

XX PN

W0200039299-A2.

QY 1669 TGGAAAGATAGCTTTCTGATAAGGAAAAGTTGCAGCTCAAGCCTATATCAATGAAGA 1728
 Db |||||
 QY 1740 TAAAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGAGCCAGGCTTATGCTTAAGAGAA 1799
 Db |||||
 QY 1729 AGGTATCTTACCTCCATCTCCAGACGACAGATGTTAAAGCAAAATCCAACTGGAGATAGTC 1788
 Db |||||
 QY 1800 AGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAATATCTGAGGCAAAAGGAGC 1859
 Db |||||
 QY 1789 AGCAGCTATTACATCTGTTGAAAGGGGAAAACGAATTCACCTGCTGCTCCATTA 1848
 Db |||||
 QY 1860 AGAAGCTATCTCAACCGCGTGAAAGCAGCTAAGAAGGTGCCACCTTGCATGCTGCTTA 1919
 Db |||||
 QY 1849 TATGTTGAGCATACATGTTGAGTGTAAAAACGGTAAATTTGATTTATCTCTCATGAAGCATCA 1908
 Db |||||
 QY 1920 CAATCTTCAATATCTGTAGAGTCAAAAACGGTAGTTTAACTATACCTCAATATGACCA 1979
 Db |||||
 QY 1909 TTACCAATAATTAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTA 1968
 Db |||||
 QY 1980 TTACCAATAACATCAAAATTTGAGTGGTTTGCAGGAAGCCCTTTATGAGGCACCTTAAGGGTA 2039
 Db |||||
 QY 1969 TACCTTGGAGATTTGTTTGGACGATTAAGTACTAGCTAGAACCCCTGACGAAGCTCC 2028
 Db |||||
 QY 2040 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATFCCAAACGAAACGCTC 2099
 Db |||||
 QY 2029 ACATTTCTAATGATGGATGGGCAATGCCAGTGAGCATGTTGTTAGGCAAGAAAGACACAG 2088
 Db |||||
 QY 2100 GCATTCAGATAATGTTTGTGTAACGCTAGCAGCAGTTCCTCAAGAAACAAAATGGTCA 2159
 Db |||||
 QY 2089 TGAAGATCCAAATAGAACTTCAAAGCGGATGAAGAG----- 2125
 Db |||||
 QY 2160 AGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAACCTTGA 2219
 Db |||||
 QY 2126 ----- 2125
 Db |||||
 QY 2220 GGAAGAAACCCCTCGAAGAGAGAAACCAAAAGCGAGAAACAGAGTCTCCAAAACCAAC 2279
 Db |||||
 QY 2126 -----CCAGTAGGAGAAACACCTCTCTGAGCGAGAAAGTCCCTCAAGTAGAGACTGAAA 2178
 Db |||||
 QY 2280 AGAGGAAACCAAGAGAGAGATCACCAGAGGAAATCAGAGAACTCAGTCCAGACTGAAA 2339
 Db |||||
 QY 2179 AGTAGAGCCCACTCAAAGAGCAGAAAGTTTGTGCTGCGAAAGTAACGGATTTCTAGTCT 2238
 Db |||||
 QY 2340 GGTTCGAAGAAAACCTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGATCCAATTAT 2399
 Db |||||
 QY 2239 GAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAATAATTTGACTCTTCAATATAT 2298
 Db |||||
 QY 2400 CAAGTCCAAATGCCAAGAGACTCTCAGAGATTAATAAATTAATTTTGGCACCCA 2459
 Db |||||
 QY 2299 GGATAACAATAGTATCATGCGCAGAGCAGAAAAATTTACTTGGTGTGTTAAAAAGGAAGTAA 2358
 Db |||||
 QY 2460 GGACACAACTATATATGCGCAGAGCTGAAAAAATTAATTTGGCTTTTAAAGGAGATTA 2519
 Db |||||
 RESULT 10
 ID AAA65736
 XX AAA65736 standard; DNA; 2647 BP.
 AC
 XX
 DT 21-NOV-2000 (first entry)
 DE
 XX Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
 DE
 XX Streptococcus pneumoniae; BVH-3; BVH-11; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX
 FN WO200039299-A2.
 XX
 PD 06-JUL-2000.

XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 XX
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 XX
 PS Example 6; Fig 15; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
 CC antigen
 XX
 SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
 Query Match 57.5%; Score 1374.2; DB 3; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 QY 1 TCTTACGAGTTGGGACTGTATCAAGCTTAGAAGCGTTAAGGAAAATAATCGTGTTCCTA 60
 Db 104 TCGTTTGAATGAGTGTGTCATCAAGCTCAAACTGTAAAGAAAATAATCGTGTTCCTA 163
 QY 61 TATAGATGAAAAAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 120
 Db 164 TATAGATGAAAAAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 223
 QY 121 GCGTGAAGGAATCAATGCTGTGAGCAAAATCGTCAATCAAGATAACAGACCAGGCTATGTCA 180
 Db 224 GCGTGAAGGAATCAACGCGGACAAATCGTCAATCAAGATTAACGATCAAGGTTATGTGAC 283
 QY 181 TTCATATGCGGACCACTATCATTTATCAATGTAAGGTTTCTTATGACGCTATCATCAG 240
 Db 284 CTCTCATGGAGACCAATTATCTACTATAATGCGAAGGTCCCTTATGATGCCATCATCAG 343
 QY 241 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAATGA 300
 Db 344 TGAAGAGCTCTCATGAAAGATCCGAATTAATCAAGTTGAAGGATTCAGACATTTGCAATGA 403
 QY 301 GGTCAAGGTTGGATATGTTTCAAGGTAGATGGAATACTATGTTTACCTTTAAGGATGC 360
 Db 404 AATCAAGGTTGGTATGTTCATTAAGGTAAACGCTAAATACTATGTTTACCTTTAAGGATGC 463
 QY 361 TGCCCAACGCGGATTAACGTCGCTACAAAAGAGAGAAATCAATCGACAAAACAAAGACATAG 420
 Db 464 AGCTCATCGGATATATGTCGCTACAAAAGAGAAATCAATCGGCAAAAACAAAGACATAG 523
 QY 421 TCAACATCGTGAAGGTGGAACCTCCAAGAAACGATGCTGCTGCTTGGCAGCTTCGCA 480
 Db 524 TCAGCATCGTGAAGGAGGACTTCAGCAAAACGATGGTGGGTAGGCTTTGCACGTTTCA 583
 QY 481 AGGACGCTTACTACAGATGATGGTTTATCTTAAATGCTTCTGATATCATAGAGATAC 540
 Db 584 GGGAGCTTACACCACAGATGATGGTTTATCTTCAATGCACTGATATCATCATGAGATAC 643
 QY 541 TGGTATGCTTATATCGTTTCTCATGGAGATCATTAACATTAATTCCTAAGATGAGTT 600
 Db 644 GGGCGATGCTTATATCGTTTCTCATGGAGATCATTAACATTAATTCCTAAGATGAGTT 703
 QY 601 ATCAGTAGCGAGTTGGCTGCTGCGAGAGCCTTCTCTATCTGTCGAGAAATCTGTCAA 660

```
Db 704 ATCAGCTAGCGAGTGGCTGCTCGAAGAGCTTCCCTATCTGGTCGGGAAATCTGCAAA 763
Qy 661 TTCAAGAACCTATCGCGGACAAATAGCGATAACACTTTCAGAACAACTGGGTACCTTC 720
Db 764 TTTAAGAACCTATCGCGGACAAATAGCGATAACACTTCCAGAACAACTGGGTACCTTC 823
Qy 721 TGTAAAGCAATCCAGGAACCTACAATACTAACAGCAACAAACAGCAACAACTAACAGTCA 780
Db 824 TGTAAAGCAATCCAGGAACCTACAATACTAACAGCAACAAACAGCAACAACTAACAGTCA 883
Qy 781 AGCAAGTCAAAGTAATGACATTCATAGTCTCTTGAACAGCTCTTACAAATCGCTTTGAG 840
Db 884 AGCAAGTCAAAGTAATGACATTCATAGTCTCTTGAACAGCTCTTACAAATCGCTTTGAG 943
Qy 841 TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCACAAATCAAGTCCGAAC 900
Db 944 TCAACGACATGTAGATCTGATGCGCTTATTTTCGACCCAGCGCAAAATCAAGTCCGAAC 1003
Qy 901 AGCTAGAGGTGTTGCAGTGCACACGAGATCAATACCACTTCATCCCTTACTCTCAAT 960
Db 1004 CGCCAGAGGTGAGTGTCCCTCATGTTAACCATTACCACTTATCCCTTATGAACAAAT 1063
Qy 961 GTCTGAATCGGAAGACGAATCGCTCGTATATTTCCCTTCGTTATCGTTCAAAACCAATG 1020
Db 1064 GTCTGAATCGGAAGACGAATCGCTCGTATATTTCCCTTCGTTATCGTTCAAAACCAATG 1123
Qy 1021 GGTACAGATTCAGGCGCAAGCAACCAAGTCCACACCGACTCCGGAACCTAGTCAGG 1080
Db 1124 GGTACAGATTCAGGCGCAAGCAACCAAGTCCACACCGACTCCGGAACCTAGTCAGG 1183
Qy 1081 CCGCAACCTCGACCAAACTTTAAATAGACTCAA-----ATTCTTCTTTGGT 1128
Db 1184 TCCGCAACCTGCACCAATCTCAACGAGCTCCAGCAATCCAAATTCATGAGAAATGGT 1243
Qy 1129 TAGTCAGCTGGTACGAAGAGTTGGGGAAGGATATGATATTTCCCTTCGTTATCGTTCAAAACCAATG 1188
Db 1244 CAAAGAAAGCTGTTGCAAAAGTAGCGATGGTTATGTTCTTTGAGGAGAAATGGAGTTTCTCG 1303
Qy 1189 TTATGCTTTGCGAAAGATTTACCATCTGAAACTGTTTAAATACTCTGAAAGCAAGTTATC 1248
Db 1304 TTATATCCGACCAAGAACTTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACTGGC 1363
Qy 1249 AAAACAAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGA 1308
Db 1364 CAACGAGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCATCTAGTGA 1423
Qy 1309 CCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTGN 1368
Db 1424 TCGAGAATTTTACAAATAAGGCTTATGACTTACTAGCAAGAAATCCCAAGATTTTACTTGA 1483
Qy 1369 AAATAAGGTCGTAATTTCTGATTTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGA 1428
Db 1484 TAATAAGGTCGACAGATTTGATTTTGGCTTTGGATTAACCTGTTGAAACGACTCAAGGA 1543
Qy 1429 TGAATCGACTAATAAGAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTC 1488
Db 1544 TGTCTCAAGTATAAAGTCAAGTTAGTGTATGATATTTCTTCCCTTCTTAGCTCCGATTCG 1603
Qy 1489 CCATCCAGAGGACTTGGCAAAACCAATTCCAATTTGAGTATACTAGACGAGATTCG 1548
Db 1604 TCATCCAGAGCGTTTAGGAAACCAATTTGGCCAAATTTACCTACACTGATGATGAGATTC 1663
Qy 1549 TATTGCTCAATTTAGCTGATAGTATACAACTGCTAGATGGTTTACATTTTGTATGAACATGA 1608
Db 1664 AGTAGCAAGTTGCGAGCAGATACACACAGAGAGCGTTTATCTTTGATCCCTCGTGA 1723
Qy 1609 TATAATCATGATGAAGAGATGCAATATGTAACCGCTCATATGCGCCCATAGTCATGAT 1668
Db 1724 TATAACCAAGTATGAGGGGATGCTATGTAACCTCCACATATGACCCCATAGCCACTGGAT 1783
Qy 1669 TGGAAAGATAGCCTTCTGATAGGAAAGTTGCAAGTTCAGCTCAAGCCTATATCAAGAAA 1728
```

```
Db 1784 TAAAAAGATAGTTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATCTCTAAAGAGAA 1843
Qy 1729 AGGTATCTCTACCTCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGC 1788
Db 1844 AGTTTGAACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGCAAAAGGAGC 1903
Qy 1789 AGCAGCTATTTCACATCGTGTGAAAGGGGAAACGAATTTCCACTGCTGCTCGACTTCATTA 1848
Db 1904 AGAAGCTATCTACAAACCGCGTGAAGCAGCTAAAGAGGTGCCACTTGTATGCTATGCCCTA 1963
Qy 1849 TATGTTGTCGATACAGTTGAGGTAAACCGTAAATTTGATTTATCTCTCATAGGATCA 1908
Db 1964 CAATCTCAATATACCTAGTAAAGTCAAAACCGGTAGTTTAAATCATACCTCTTATGACCA 2023
Qy 1909 TTACCATATATTAATTTTCCTTTGGTTTGTATGATCAACATACAAAGCTCCAAATGGCTA 1968
Db 2024 TTACCATATCAATTTGAGTTGTTTGAAGAGGCTTTTATGAGSCACCTTAAGGGGTA 2083
Qy 1969 TACCTTGGAAAGATTTGTTTCGACGATTAAGTACTAGTATAGAACACCTTCAGCAACGCTCC 2028
Db 2084 TACTCTTGGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAACGCTCC 2143
Qy 2029 ACATTTCTAATGATGGATGGGCAATCCAGTGAAGCATGTTGTAGGCAAGAAAGACACAG 2088
Db 2144 GCATTCAGATAATGGTTTTGTTAACTGCTAGCAGCATGTTTCAAGAAACAAATATGCTCA 2203
Qy 2089 TGAAGATCCAAATAAGAACTTTCAAAGCGGATGAAGAG----- 2125
Db 2204 AGCTGATACCAATCAACGGAAACCAACGCGAGGAGAAACCTCAGACAGAAAAACCTGA 2263
Qy 2126 ----- 2125
Db 2264 GGAAGAAACCCCTCGAGAGAGAGAAACCAAAAGCGAGAAACCGAGTCTCCAAACCAAC 2323
Qy 2126 -----CCAGTAGAGAAACACCTGCTGAGCGAGAAAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2324 AGAGGAACCAAGAGAGATCCACGAGGAATCAGAGAACTCAGTCCGAGACTGAAAA 2383
Qy 2179 AGTAGAGCCCAACTCAAAGAACGAGAGTTTTCGTTGCGAAAGTAAACGAGTCTAGTCT 2238
Db 2384 GGTGAGAAACAACTGAGAGAGGCTGAAGATTTACTTTGGAATAATCCAGGATCCAAATAT 2443
Qy 2239 GAAAGCAATGCAACAGAACTCTAGCTGTTTACGAAATAATTTGACTCTTCAAAATAT 2298
Db 2444 CAAGTCCAAATGCCAAAGAGACTCTCACAGGATTAATAAATAATTTACTATTGCGACCCA 2503
Qy 2299 GGATAACAAATAGTATCATGCGAAGCAGAAAAATTTACTTGCCTTGTTTAAAGGAAGTAA 2358
Db 2504 GGACAACAACTACTATTATGCGAGAGGCTGAAAACTATTGCTTTATTAAAGGAGAGTAA 2563

RESULT 11
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX
AC ABK15103;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
streptococcal bacterial infection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
PH Key Location/Qualifiers
FT CDS 45..2567
FT /product= a
FT /note= "BVH-11"
FT FT SP64, no information on which is given in the
FT FT specification
```


1429 TGAATCGACTAATAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACAATTAC 1488
 1544 TGCTCAAGTGATAAAGTCAAGTTAGTGAATGATTTCTTGGCTTCTTAGCTCGATCG 1603
 1489 CCATCCAGAGCGACTTTGGCAAAACCAATTTCTAAATTTAGTATATCTGAAGACGAGTTG 1548
 1604 TCATCCAGAGCGTTTGAAGAAAAACCAATGCGCAATTTACCTACACTGATGATGATTCA 1663
 1549 TATTGCTCAATTAGCTGATAGTATACAACTGATGATGGTTTACATTTTGTGATGAACATGA 1608
 1664 AGTAGCCCAAGTTGGCAGCAGATGATACACACAGAGACGGTTATATCTTTGATCCTCGTA 1723
 1609 TATAATCAGTCATGAAGAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGAT 1668
 1724 TATAACCAAGTGTAGGGGAGTCTATGTAATCTCCACATATGACCATAGCCACTGGAT 1783
 1669 TGGAAAAAGATAGCCCTTTCTGTATAGGAAAAAGTTGCGAGCTCAAGCTTATACCTAAGAAAA 1728
 1784 TAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCGACCCAGGCTTATGCTAAAGAGAA 1843
 1729 AGGTATCCTACCTCCATCTCCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTC 1788
 1844 AGTTTACGCCCTCTTCGACAGCAATCAGGATTCAGGAAATATCTGAGGCAAAAGAGC 1903
 1789 AGCAGCTATTATCAATCGTGTGAAGGGGAAAAACGAATTCACACTCGTTGACCTTCCATA 1848
 1904 AGAAGCTATCTACAAACCGGTGAAGCAGCTAAGAGGTGCCACTTGTATGCTATGCCCTTA 1963
 1849 TATGGTTGAGCATACAGTTGAGGTTAAACCGTAATTTGATTTATCTCTATAGGATCA 1908
 1964 CAATCTTCAATATCTGTAGAAGTCAAAAACGGTAGTTTATCATCTCATCTATTATGACCA 2023
 1909 TTACCAATATTAATTAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTA 1968
 2024 TTACCAATATTAATTAATTTGAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 2083
 1969 TACCTTGGAGATTTGTTGGACGATTAAGTACTAGTACGACACACCTGACGACGTC 2028
 2084 TACTCTTGGAGATCTTTGGGAGCTGTCAAGTACTATGCGAACATCCAAACGACGTC 2143
 2029 ACATTCTAATGATGATGGGCAATGCGAGTGAGCATGTTGTAGCAAGAAGACCAAG 2088
 2144 GCATTTCAGATAATGTTTGTGTAAGCTAGGACCATGTTTCAAAAGAAACAAAAATGCTCA 2203
 2089 TGAAGTCCCAATTAAGACTTCAAGCGGATGAAG 2125
 2204 AGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCGACAGAAAAACCTGA 2263
 2126 ----- 2125
 2264 GGAAGAAACCCCTCGAGAAGAGAAACCAACCAAGCGAGAAACAGAGTCTCCAAAACCAAC 2323
 2126 -----CCAGTAGAGGAAACACCTGCTGAGCCAGAGTCTCCTCAAGTAGAGACTGAAA 2176
 2324 AGAGAAACCAAGAGAAATCACCAGAGGAATCAGAGAACTCAGGTCGAGACTGAAAA 2383
 2179 AGTAGAACCCCACTCAAGAGAGAGAGTTTGTCTTGGAAAGTAACGGATTTCTAGTCT 2238
 2384 GGTGAGAAACCACTGAGAGGCTGAGATTTACTTGGAAATCCAGGATCCAAATAT 2443
 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATATTTGACTCTTTCAATAT 2298
 2444 CAAGTCCCAATGCAAGAGACTCTCACAGGATTAAGAAATATTTACTATTGTCACCCA 2503
 2299 GGAATCAATAGTATCATGCGAGAGCAGAAAAATTTCTGGTTGTTTAAAGGAGTAA 2358
 2504 GGACAAACATATATTATGAGAGAGCTGAAAAACCTATTGGCTTTATTAAAGAGAGTAA 2563

RESULT 12
 AAA08557
 ID AAA08557 standard; DNA; 2478 BP.
 XX

AAA08557;
 19-JUL-2000 (first entry)
 S. pneumoniae 92 kDa human C3-degrading protein coding sequence.
 Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
 inhibitor; inflammation; organ rejection; xenotransplantation; ss.
 Streptococcus pneumoniae.
 W0200017370-AL.
 30-MAR-2000.
 24-SEP-1999; 99WO-US022362.
 24-SEP-1998; 98US-0101736P.
 31-MAR-1999; 99US-00283094.
 (MINU) UNIV MINNESOTA.
 (AMCY) AMERICAN CYANAMID CO.
 Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
 WPI; 2000-283594/24.
 P-PSDB; AAY91939.
 Isolated polypeptide is used to stimulate immune system and immunize or
 treat a mammalian subject against Streptococcus pneumoniae infection or
 colonization.
 Claim 1; Page 55-57; 63pp; English.
 The present sequence, isolated from Streptococcus pneumoniae, encodes a
 human C3-degrading protein (see AAY91939) of about 92 kDa. This sequence
 may encompass a smaller 20 kDa polypeptide coding sequence (AAA08556)
 also having human C3-degrading activity. The DNA sequences can be used
 for producing an immune response to Streptococcus pneumoniae in a mammal.
 Antibodies against the proteins can be used to inhibit S. pneumoniae.
 mediated C3 degradation. C3-mediated inflammation and rejection in
 xenotransplantation can be inhibited by expressing the nucleic acid
 sequences on the surface of an organ of an animal. In particular, the
 polypeptides are useful for stimulating the immune system and are
 effective to immunize or treat a mammalian subject against Streptococcus
 pneumoniae infection or colonization
 Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 U; 0 Other;
 Query Match 42.4%; Score 1011.8; DB 3; Length 2478;
 Best Local Similarity 66.2%; Pred. No. 3.8e-245;
 Matches 1628; Conservative 0; Mismatches 688; Indels 144; Gaps 6;
 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTC 57
 60 TTCTTATGAATCTGCTGTCCACAGCTGGTCAGTTAAGAGAGTCTAATCGATTTC 119
 58 CTATATAGATGGAAAAAAGCGAGCGACGCAAAAACGGAGAAATTTGACTCTGTAGAGTTAG 117
 120 TTATATAGTGTGATCAGGCTGGTCAAAAGGCGAGAAAACCTTGACACAGATGAAGTCAG 179
 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCTCAAGATAACAGACCAAGGCTATGT 177
 180 TAAGAGGAGGGGATCAACGCGCAACAATCGTCTCAAGATTAAGGATTCGGATCAAGGTTATGT 239
 178 CACTTCACATGGGACCACTATCATTAACAATGGTAAGGTTCTTATGACGCTATCAT 237
 240 GACCTCTCATGAGACCACTTATCATTAATTAAGGAGGCTCCCTTATGATGCCATCAT 299
 238 CAGTGAAGAATTTACTCATGAAGAATCCAAACTATAGCTTAAAGATGAGGATATTGTTAA 297
 300 CAGTGAAGAGCTCCTCATGAAGAATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAA 359

QY 298 TGAGGTCAAGGTTGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGA 357
 Db 360 TGAATCAAGGTTGGTATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGA 419
 QY 358 TGCTGCCACGCGGATTAACGTCGTTCAAAAGAGGAAATCAATTCGACAAAACAAAGAGCA 417
 Db 420 TGCAGCTCATCGGATAATATTCGGACAAAAGAGAGATTAAACGTCAGAAGCAGGAACA 479
 QY 418 TAGTCACATCGTGAAGGTGGAACTCAAGAAACGATGGTGTGCTGTCCTTGGCCTTGGCAGCTTC 477
 Db 480 CAGTCATAATCAACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGACCCAGAGC 533
 QY 478 GCAAGGACGCTTACTACAGATGATGGTTATATCTTTTAATGCTTCGATATCATAGAGGA 537
 Db 534 CCAAGGACGCTATACACCGATGATGGTTATATCTTCAATGCACTGATATCATTTAGGGA 593
 QY 538 TACTGGTATGCTTATATATCGTTCTCATGGAGATCAATTACCAATTACATTCCTTAAGGAATGA 597
 Db 594 CACGGGTGATGCTTATATCGTTCTCACGGCGACCAATTACCAATTACATTCCTTAAGGAATGA 653
 QY 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCCTATCTGGTCGAGGAATCTGTC 657
 Db 654 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAGCCTTATGGAAATGG----- 698
 QY 658 AAATTCAAGAACTATCGCGCGCAAAATAGCGAATAACACTTCAAGAAACAAACTGGGTACC 717
 Db 699 -GAAGCAGGATCTCGTCTCTTCAAGTTCCTAGTTTATGCAATTCAGCTCAACCAA 757
 QY 718 TTCTGTGAAGCAATCCAGGAACTACAAATATAACACAGCAACACAGCAACACTAACAG 777
 Db 759 GATTGTGAGAGAACCAATCT-----GACTGTCACTCCAACTTA 797
 QY 778 TCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAACAGCTCTACAAACTGCTTT 837
 Db 798 TCATCAAAATCAAGGGGAAAAATTTCAAGCCCTTTTACGTGAATGTTATGCTTAACCCCTT 857
 QY 838 GAGTCAACGACATGTAGAACTCTGATGCCCTTGCTTTGATTCAGACACAAATCACAGTCG 897
 Db 858 ATCAGNACGCCATGTGGAACTGTAGGCCCTTATTTTCGACCACGCCCAATCACAAAGTCG 917
 QY 898 AACAGCTAGAGGTGTGCAAGTGCCACACGAGATCAATTACCACTTCATCCCTTACTCTCA 957
 Db 918 AACCGCCAGAGGTGTAGCTGTCCTCATGTAAACCAATTACCACTTTATCCCTTATGAACA 977
 QY 958 AATGTCGAATTTGGAAGACGAATCGCTGTTATTTATCCCTTCGTTATCGTTTCAAAACCA 1017
 Db 978 AATGTCGAATTTGGAAGAACGAATGCTCGTATTTATCCCTTCGTTATCGTTTCAAAACCA 1037
 QY 1018 TTGGGTACCAGATTCAAGGCCAGAACCAACAACTCCACACCGACTCCGGAACCTAGTCC 1077
 Db 1038 TTGGGTACCAGATTCAAGACCAGAACCAACAACTCCACAGCTCCGGAACCTAGTCC 1097
 QY 1078 AGGCCGGAACTGCAACCAATCTTAAATAAGACTCAA-----ATTCTTCTTT 1125
 Db 1098 AAGTCGGCAACTCGCACCAAACTCTCAACAGCTCCAGCAATCCCAATTTGATGAGAAAT 1157
 QY 1126 GGTAGTCAAGTGTGAGAAAGTTGGGAAAGGATATGTTTTCGAAGAAAGGCAATCTC 1185
 Db 1158 GGTCAAGAGAGCTGTTTCGAAAAGTAGCGGATGGTTATGTTTGGAGGAAGTGGAGTTTC 1217
 QY 1186 TCGTTATGTCCTTGGCAAGATTTACCATCTGAACTGTTAAAAATCTTTGAAAGCAAGTT 1245
 Db 1218 TCGTTATATCCAGCCAGGAATCTTTCAGCAGNAAACAGCAGCAGGCAATTGATAGCAACT 1277
 QY 1246 ATCAAAACAGAGAGTGTTCACACTTTAACTGCTAAAAAGAAAAATGTTGCTCCTCG 1305
 Db 1278 GGCCAAAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAAACTGACCTCCCACTAG 1337
 QY 1306 TGACCAAGAAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
 Db 1338 TGATCGAGAAATTTTACAATAAGCTTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
 QY 1366 TGNAAATAAGGTCGTAATTTCTGATTTTCAAGCCTTTAGACAAATATTATAGACGCTTGAA 1425

Db 1398 TGATAATAAAGTTCGACAAAGTTGATTTTGAGGCTTTGGATTAACCTGTTGGAACGACTCAA 1457
 QY 1426 TGATGAATCGACTAATAAAGAAAAATTTGGTAGATGATTTTATTGGCAATCTCTAGCACCAT 1485
 Db 1458 GGATGTCCTCAAGTGTATAAAGTCAAGTTAGTGGATGATATCTTGGCCTTCTTAGCTCGAT 1517
 QY 1486 TACCCATCCAGAGCGACTTTGGCAACCAATTTCTCAATTTAGTATATCTAGTGAAGCAAGT 1545
 Db 1518 TCGTCATCCAGAACGTTTAGGAAAAACCAATCGCAATTTACCTACACTGATGATGAGAT 1577
 QY 1546 TCGTATTTGCTCAATTAGCTGATAAGTATACAAGTCAGATGGTTACATTTTTGATGAACA 1605
 Db 1578 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAACGGTTATATCTTTGATCCTCG 1637
 QY 1606 TGATATAATCAGTGTGAAGGAGATGCATATGTAAAGCGCTCATATGGGCCATAGTCACTG 1665
 Db 1638 TGATATAACCAAGTGTAGGGGGATGCTTATGTAACTCCACATATGACCCATAGCCACTG 1697
 QY 1666 GATTGGMAAGATAGCCTTTCTGTATAAGAAAAAGTTGCAGCTCAAGCCTATCTACTAAGA 1725
 Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCGCAGCCAGGCTTATGCTAAAGA 1757
 QY 1726 AAAAGTATCTTACCTCCATCTCCAGACGACATGTTAAAGCAAAATCCAACTGGAGATAG 1785
 Db 1758 GAAAGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGAAATACTAGAGCAAAAG 1817
 QY 1786 TCGAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAAGAAATTCCACTCGTTTCGACTTCC 1845
 Db 1818 AGCAGAAAGCTATCTCAACCGCTGAAAGCAGTAAAGAGGTGCCACTTTGATCGTATGCC 1877
 QY 1846 ATATATGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTTATTTCTCATAGA 1905
 Db 1878 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAACTCATACCTCATATGA 1937
 QY 1906 TCATTACCAATAATATAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGG 1965
 Db 1938 CCAATTACCAATAACATAAATTTGAGTGGTTTGAAGAGCCTTTATGAGGACCTTAAGG 1997
 QY 1966 CTATACCTTTGGAAGATTTGTTTGCAGCAGTAAAGTACTACGTAGAACACCTTGACGAAGC 2025
 Db 1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCCAACGAAGC 2057
 QY 2026 TCCACATCTTAATGATGGGCAATGCCAGTAGAGCTGTTAGGCAAGAAAGACCA 2085
 Db 2058 TCCGCAATTCAGATAATGGTTTGGTAAACGCTAGCAGCAATGTTCAAGAAAGAAACAAAATGG 2117
 QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGA----- 2121
 Db 2118 TCAAGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
 QY 2122 ----- 2121
 Db 2178 TGAGGAAGAAACCCCTCGAGAGAGAAACCGAAAGCGAGAAACCCAGAGTCTCCAAAACC 2237
 QY 2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAANA 2178
 Db 2238 AACAGAGAACCCAGAGAAATCCAGAGGAATTCAGAAAGAACCTCAGGTCGAGACTGAAA 2297
 QY 2179 AGTAGAAGCCCAACTCAAAAGAAAGAGAGTTTGTCTCGAAAGTAAACGGATTTCTAGTCT 2238
 Db 2298 GGTTCAGAAAAACTGAGAGAGCTGAGATTTACTTGGAAAAATCCAGGATCCAAATTAT 2357
 QY 2239 GAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAATTTGACTTTCAAATTAT 2298
 Db 2358 CAAGTCCAAATGCAAAAGAGACTCTCACAGGATTTAAAAAATAATTTACTATTTGGCACCCA 2417
 QY 2299 GATTAACATAGTATCATGGCAGAGACGAGAAAAATTTACTTGGTTGTTTAAAAAGGAGTAA 2358
 Db 2418 GGACAACAATACTATTATGGCAGAGAGCTGAAAAACATTTATGGCTTTATTTAAAGAGAGTAA 2477

AAA05417	QY	238	CAGTGAAGAATTA	297
ID AAA05417 standard; DNA; 2481 BP.				
XX	Db	300	CAGTGAAGAGCT	359
AC AAA05417;				
XX	QY	298	TGAGGTCAAGGGT	357
DT 24-MAY-2000 (first entry)				
XX	Db	360	TGAAATCAAGGGT	419
DE Streptococcus pneumoniae nucleotide sequence ID311.				
XX	QY	358	TGCTGCCACGCGA	417
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;				
KW antibacterial; antinflammatory; meningitis; infection; diagnosis;				
KW pneumococcal disease; ds.				
XX	Db	420	TGAGGTCAAGGGT	479
OS Streptococcus pneumoniae.				
XX	QY	418	TAGTCAACATCGT	477
XX	Db	480	CAGTCATAATCAC	533
PN W0200006737-A2.				
XX	QY	478	GCAGGACGCTATA	537
PD 10-FEB-2000.				
XX	Db	534	CCAAGGACGCTAT	593
PF 27-JUL-1999; 99WO-GB002451.				
XX	QY	538	TACTGGTGAATCT	597
PR 27-JUL-1998; 98GB-00016337.				
PR 19-MAR-1999; 99US-0125164P.				
XX	Db	594	CACGGGTGATGCT	653
XX (MICR-) MICROBIAL TECHNIQS LTD.				
PA Gilbert CFG, Hansbro PM;				
XX	QY	598	GTTATCAGCTAGC	657
PI	Db	654	GTTATCAGCTAGC	698
XX	QY	658	AAATTCAAGAACCT	717
DR WPI; 2000-195300/17.				
DR P-PSDB; AAY81662.				
XX	Db	699	-GAAGCAGGGATCT	757
XX New Streptococcal protein, useful as a vaccine, for diagnosis of				
PT pneumococcal diseases and for screening agents capable of antagonizing or				
PT inhibiting expression of the protein.				
XX	Db	758	GATTGTCAGAGACC	797
PS Claim 2; Page 99; 108pp; English.				
XX	QY	778	TCAGCAAGTCAAAG	837
CC	Db	798	TCATCAAAATCAAG	857
CC	QY	838	GAGTCACGACATGT	897
CC	Db	858	ATCAGAACGCGCAT	917
CC	QY	898	AAACAGCTAGAGGT	957
CC	Db	918	AAACGCGCAGAGGT	977
CC	QY	958	AATGTCCTGAATT	1017
CC	Db	978	AATGTCCTGAATT	1037
SQ	QY	1018	TTGGGTACCGAGTTC	1077
Query Match 42.0%; Score 1003.8; DB 3; Length 2481;				
Best Local Similarity 66.0%; Pred. No. 4e-243;				
Matches 1623; Conservative 0; Mismatches 693; Indels 144; Gaps 6;				
QY	1	TTCTACGAGTTGGAC	57	TTATCTGTTTC
Db	60	TTCTATGAGCTTGGT	119	TTATCTGAGTTGC
QY	58	CTATATAGATGGA	117	TTATCTGAGTTAG
Db	120	TTATATAGATGTT	179	TTATCTGAGTTAG
QY	118	CAACGCTGAGGAAT	177	TTATCTGAGTTAG
Db	180	TAAGAGGAGGGGAT	239	TTATCTGAGTTAG
QY	178	CACCTTCATCGGAC	237	TTATCTGAGTTAG
Db	240	GACCTTCATCGGAC	299	TTATCTGAGTTAG

Db 1338 TGATCGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGATTCACCAAGATTACT 1397
 QY 1366 TGNAAATAGGGTGTGAATTCGTATTTTCCAGCTTAGCAAAATTAATAGACGGTTGAA 1425
 Db 1398 TGATAATAAGGTCGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAACGACTCAA 1457
 QY 1426 TGATGAATCGACTTAATAAGAAAAATTTGTAGATGATTTATTGGGATTCCTAGCACCAAT 1485
 Db 1458 GGATGTCCTCAAGTGATAAAGTCAAGTTAGTGATGATATCTTGGCTTCTTAGCTCCGAT 1517
 QY 1486 TACCATCCAGACGACTTTGGCAACCAAAATTTCTCAAATTTGAGTATPACTGAAGACGAAGT 1545
 Db 1518 TCGTCATCCAGAACGTTTAGGAAAAACCAATGCGCAAAATTAACCTACACTGATGAGAT 1577
 QY 1546 TCGTATTGCTCAATAGCTGATAAGTATACAGCTGAGATGTTTACATTTTGTATGAACA 1605
 Db 1578 TCAAGTAGCCTAAGTTGGCAGGCAAGTACACAGAAAGCGGTTATATCTTTGATCCTCG 1637
 QY 1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTG 1665
 Db 1638 TGATATAACCAAGTATGAGGGGATGCCCTATGTAACCTCACATATGACCATAGCCACTG 1697
 QY 1666 GATTGGAAAGATAGCCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATATAAAGA 1725
 Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGGCGGAGCCAGGCTTTATGCTAAAGA 1757
 QY 1726 AAAAGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAATCAACTGGAGATAG 1785
 Db 1758 GAAAGGTTTGACCCCTCTTCCAGACACCATCAGGATTCAGGAAATPACTGAGGCAAAAG 1817
 QY 1786 TCAGCAGCTATTATCAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTTCGACTCC 1845
 Db 1818 AGCAGAGCTATCTACACCGGCTGAAACAGCTAAGAAGGTGCCACTTGTATGCTATGCC 1877
 QY 1846 ATATATGTTGAGCTACATAGTTGAGGTTAAACCGGTAAATTTGATTAATTCCTATAAGA 1905
 Db 1878 TTACAATCTTCAATATCTATGTAAGTCAAAACCGGTAGTTTAAATCATACCTCATATGA 1937
 QY 1906 TCATTACCATATATAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGG 1965
 Db 1938 CATTACCATACATCAATTTTGAAGTGTGTTGACGAGGCTTTTATGAGGCACTTAAGG 1997
 QY 1966 CTATACCTTGGAGATTTCTTTGCGACGATTAAGTACTAGTAGAACACCTGAGCAAGC 2025
 Db 1998 GTATCTCTGAGGATCTTTTGGGACTGTCAAGTACTATGTGCAACATCTCAACCAAGC 2057
 QY 2026 TCCATCTTAATGATGATGGGCAATCCAGTGAGCATGTTGTTAGGCAAGAAAGACCA 2085
 Db 2058 TCCGCAATCAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAGAAACCAAAAATGG 2117
 QY 2086 CAGTGAAGTCAATTAAGAACTTCAAGCGGATGA----- 2121
 Db 2118 TCAAGCTGATACCAATCAAAACGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
 QY 2122 ----- 2121
 Db 2178 TGAGGAGAAACCCCTCGAGAGAGAGAAACCGCAAGCGAGAAACACAGAGTCTCCAAACC 2237
 QY 2122 ---AGAGCCAGTAGAGAAACACCTGCTGAGCGAGAGTCCCTCAAGTAGAGACTGA AAA 2178
 Db 2238 AACAGAGGAACCAAGAAATCACCAGAGGAATCAGAGAAACCTCAGGTCGAGACTGAAA 2297
 QY 2179 ACTAGAGCCCACTCAAGAGAGCAGAGATTTTGGCTTGGCAAGTAACGATTTAGTCT 2238
 Db 2298 GGTGGAAGAAACCTGAGAGAGGCTGAAGATTTACTTTGAAAAATCCAGAGTCCAAATTAT 2357
 QY 2239 GAAAGCAATGCAACAGAACTCTAGCTGTTTACGAAATATTTGACTCTTCAATTAT 2298
 Db 2358 CAAGTCCATGCAAGAGACTCTACAGATTTAAAAATTAATTTACTATTTGGCACCCA 2417
 QY 2299 GGATAACAATAGTATCATCGCAGAGACAGAAAAATTTACTTGGCTTGTAAAAAGGAGTAA 2358

Db 2418 GGACAACRATACTATTATGGCAGAGAGCTGAAAAACTATTGGCTTTATTAAAGAGAGTAA 2477
 RESULT 14
 ABX06885
 ID ABX06885 standard; DNA; 2457 BP.
 XX
 AC ABX06885;
 XX
 DT 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 S. pneumoniae type 4 strain coding region #1173.
 XX
 Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 ear infection; antiinflammatory; antibacterial; immunostimulant;
 auditory; respiratory; gene therapy; vaccine.
 XX
 Streptococcus pneumoniae; type 4 strain.
 OS
 WO200277021-A2.
 XX
 PN 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-IB002163.
 XX
 PR 27-MAR-2001; 2001GB-00007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR P-PSDB; ABU01597.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 useful as medicaments for treating or preventing a disease or infection
 due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 ear infection.
 PT
 XX
 PS Claim 6; SEQ ID NO 2345; 56pp; English.
 CC
 CC The invention relates to a protein comprising or having at least 50%
 identity to any of the 2469 amino acid sequences, identified in the
 specification (available on a computer readable format), or its fragment,
 expressed from 2469 of 2489 identified DNA coding regions from the
 Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 ABS56454. Also included are an antibody which binds one of the proteins,
 treating a patient by administering the protein, DNA or antibody (in a
 composition), a kit comprising first and second primers, which are the
 nucleic acid cited above or fragments between nucleotides 8-100 of a
 sequence not defined in the specification, for amplifying a target
 sequence contained within a Streptococcus nucleic acid sequence, where
 the first primer is substantially complementary to the target sequence
 and the second primer is substantially complementary to the complement of
 the target sequence, and where the parts of the primers having
 substantial complementarity define the termini of the target sequence to
 be amplified, assay comprising contacting a test compound with the
 protein, and determining whether the test compound binds to the protein
 and a Streptococcus pneumoniae bacterium, where one or more genes
 encoding the proteins has been rendered inactive. The proteins, nucleic
 acid molecules, antibody and compositions are useful as medicaments for
 treating or preventing a disease or infection due to streptococcus
 bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 media or ear infection. They are also useful in developing vaccines,
 diagnostics and antibiotics. The methods are useful for identifying
 immunodominant proteins. The present sequence is one of the 2489
 identified coding region from the genomic sequence. Note: The sequence
 data for this patent did not form part of the printed specification, but
 was obtained in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 standardise OS field)

XX	Sequence	2457 BP; 836 A; 497 C; 531 G; 593 T; 0 U; 0 Other;	
SQ	Query Match	41.6%; Score 993; DB 7; Length 2457;	
	Best Local Similarity	65.8%; Pred. No. 2.1e-240;	
	Matches 1611; Conservative	0; Mismatches 696; Indels 141; Gaps 6;	
Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC	57
Db	60	TTCTTATGAGCTTGGAGCTTACCAAGCTGGTCAGGATAAGAAAGTCTAATCGAGTTGC	119
Qy	58	CTATATAGATGGAAAAACAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTAG	117
Db	120	TTATATAGATGGTGATCAGGCTGTCAAAGGCGAGAAAACCTTGACACAGATGAATCAG	179
Qy	118	CAAGCGTAAGGAATCAATGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGCTATGT	177
Db	180	TAAGAGGAGGGGATCAACGCGAACAATTTGTTATCAAGATTACGGATCAAGGTTATGT	239
Qy	178	CACCTTCAATGGGACCACTATCATTTATTAATGATGTAAGGTTCCCTTATGACGCTATCAT	237
Db	240	GACCTTCAATGGAGACCAATTAATTAATTAATGAGCAAGGTTCCCTTATGATGCCATCAT	299
Qy	238	CAGTGAAGAAATTAATCAATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTAATCAGTTGAAGGATTCAGACATTTGCAA	359
Qy	298	TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA	357
Db	360	TGAATCAAGGGTGGTATGTCATTAAGGTAACCGTAAATTAATGTTTACCTTAAGGA	419
Qy	358	TGCTGCCACGCGATAAAGCTGCTGACAAAGAGAAATCAATCGACAAAACCAAGAGCA	417
Db	420	TGAGCTCATGCGGNTAATATTCGACAAAAGAGAGATTAACGTCAAGACGAGAAAG	479
Qy	418	TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGGTGTCTGTCCTTGGCACGTTTC	477
Db	480	CAGTCATAATCAT-----AACTCAAGACGAGATAATGCTGTGCTGCAGCCAGAGC	530
Qy	478	GCAGGACGCTATACTACAGATGATGTTATATCTTAATGCTTCTCATATCATAGAGGA	537
Db	531	CCAAAGGAGCTTATCAACGAGATGATGGTATATCTTCAATGCACTCATATCATAGAGGA	590
Qy	538	TACTGTGATGCTTATATCGTCTCTCATGGAGATCAATACCATACATTCCTAAGAATGA	597
Db	591	CACGGGTGATGCTTATATCGTCTCTCACGGGACCAATACCATTTACATTCCTAAGATGA	650
Qy	598	GTTATCAGCTAGCGAGTTGGTGTGCGAGAGCCTTCTCATCTGGTCGAGGAAATCTGTC	657
Db	651	GTTATCAGCTAGCGAGTTAGTGTGCTGCAGAGCCCTATTGGAAATGG-----	695
Qy	658	AAATTCAGAACCTTATCGCGACAAAATAGCGATAACACTTCAAGAACCAACTGGGTACC	717
Db	696	-GAAGCGGAGATCTCGCTCTTCTTCAAGTCTAGTTATATGTAATGCAAACTCCAGCTCAACCAA	754
Qy	718	TTCTGTAAAGCAATCCAGGAACTACAAATTAATCAACAGCAACCAACAGCAACATAACAG	777
Db	755	GATTTGTCAGAGAACCAATCT-----GACTGTCACTCCCACTTA	794
Qy	778	TCAAGCAAGTCAAAAGTAAATGATGATGATGCTCTTGAACAGCTCTACAACTGCCTTT	837
Db	795	TCATCAAAATCAAGGGGAAAACATTTCAAGGCTTTTACGTGAATTTGATGCTAAACCCCTT	854
Qy	838	GAGTCAACGACATGTAGAATCTGATGCTGTCTTTGATGTCAGCACAAATCACAAGTCG	897
Db	855	ATCAGAACGCAATGTGGAATCTGATGGCTTATTTTTCAGCCAGCGCAATCAACAGTCG	914
Qy	898	AACAGCTAGAGGTGTGAGTGCCACACGAGATCATTAACACTTCACTCCCTTACTCTCA	957
Db	915	AACGCCAGAGGTGTAGTGTCTCTCATGTTAAACCAATTAACCACTTATATCCCTTATGAACA	974
Qy	958	AATGCTGAATGGAAGAACCAATCGCTCGTATTAATTTCCCTTCTGTTATCGTTCAAAACCA	1017

Db	975	AATGCTCTGAATTTGAAAAACGAATTTGCTCGTATTAATTTCCCTTCTGTTATCGTTCAAAACCA	1034
Qy	1018	TTGGGTACAGATTTCAAGGCCAGAACAAACCAAGTCCACACCGACTCCGGAACCTACTGCC	1077
Db	1035	TTGGGTACAGATTTCAAGAACAGAGAACCAAGTCCACACCGACTCCGGAACCTACTGCC	1094
Qy	1078	AGGCCCGCAACCTGCACCAATCTTAAATATAGACTCAAAATCTTCTTTTGGTTAGTCAAGCT	1137
Db	1095	AAGTCCGCAACAGCTCCAAGCAATCCAATGTATGAGAA-----ATTGTCRAAGAGC	1148
Qy	1138	GGTACGAAAAAGTTGGGGAAGGATATGATTTTCGAAGAAAAAGGCAATCTCTCTGTATGTCCT	1197
Db	1149	TGTTCCAAAAAGTAGGCGATGTTATGTTTTCGAGGAGATGAGTCTTCTCGTTATATATCC	1208
Qy	1198	TGCGAAGATTTACCATCTGAAACTGTTAAATCTTTGAAGCAAGTTATCAAAACAGA	1257
Db	1209	AGCCAAAGGATCTTTTCAGCAAAAACAGCAGCGCAATTTGATAGCAAACTGGCCRAAGCAGA	1268
Qy	1258	GAGTGTTCACACACATTTTAACTGCTAAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAT	1317
Db	1269	AAGTTTATCTCATAAAGCTAGGAACCTAAGAAAACTGACCTCCCATCTAGTGATCGAGAAT	1328
Qy	1318	TTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATATAGGG	1377
Db	1329	TTCAATATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTTGATATAAAGG	1388
Qy	1378	TCGTAAATCTGATTTTCCAAAGCTTAGACAAATTAATTAAGACGCTTGAATGATGAATCGAC	1437
Db	1389	TCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAAAGCACTCAAGATGTCCTAAG	1448
Qy	1438	TAATAAAGAAAAATGTTGATAGATGATTTTATTTGGCATTCCTAGCACCAATTAACCCATCCAGA	1497
Db	1449	TGATAAAGTCAAGTTAGTGAAGATATTTCTTGCCTCTTAGCTCCGATTCGTATCCAGA	1508
Qy	1498	CGACTTGGCAAAACCAAAATTTCAAATTTAGTATCTAGTGAAGCAAGTTCGTATGTCCTCA	1557
Db	1509	ACGTTTAGGAAAAACCAATTCGCAAAATTTACTCTACATGATGATGAGATTTCAAGTAGGCAA	1568
Qy	1558	ATTAGCTGTAAGTATACAAACGTCAGATGGTTACATTTTGTGAACATGATATAATCAG	1617
Db	1569	GTTGGAGGCAAGTACAAACAGAGACGGTTATATCTTTGATCTCTGATATAACAG	1628
Qy	1618	TGATGAAGAGATGCAATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAATCCACATATGACCCATAGCCACTGGATTTAAAAAGA	1688
Qy	1678	TAGCTTCTTGATGAAGAAAAAGTTGCACTCAAGCCTATATCAAGAAAAAGGATATCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGCGCGACGCCAGGCTTATGCTTAAAGAGAAAGGTTTGAAC	1748
Qy	1738	ACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTAT	1797
Db	1749	CCCTCTTCAGACAGCATCAGGATTCAGGAAATACTAGGCAAAAGAGAGCAGAGCTAT	1808
Qy	1798	TTACAATCTGTGTGAAGGGGAAAAACGAATTTCCATCTCGTTCGACTTCCATATATGTTGA	1857
Db	1809	CTACAACCGCGTGAAGCAGCTAAGAAAGTGCCACTTGATGCTATGCTTACAACTTCTCA	1868
Qy	1858	GCATACAGTTGAGTTTAAACCGTAAATTTGATTTTCTCTAATAGATCATTTACCATAA	1917
Db	1869	ATATCTGTGAAGTCAAAACCGTAGTTTAACTCATACCTCATTTATGACCATTAACCATAA	1928
Qy	1918	TATTAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGA	1977
Db	1929	CATCAATTTGAGTTGTTTGAAGGCGCTTTTATGAGGCACTTAAGGGGTATCTCTTGA	1988
Qy	1978	AGATTTGTTGGAGAGTTAAGTACTAGTAGAACAACCTTGAAGAGCTCCACATTTCTAA	2037
Db	1989	GGATCTTTTGGCGACTCTCAAGTACTATGTGCAACATCCAAACGAACTCCGCAATTCAGA	2048
Qy	2038	TGATGATGGGGCAATCCCAAGTGAAGTGTGTTAGGCAAGAAAGACACAGTGAAGATCC	2097
Db	2049	TAATGGTTTGGTAAACGCTAGCGCAATGTTCAAGAAACAAAAATGGTCAAGCTGATAC	2108

QY 2098 AATAAGAACTTCAAGCGGATGA----- 2121
 Db 2109 CAATCAACGGAAGAAACCAAGCGGAGGAGAAACCTCAGACAGAAACCTGAGGAGAAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCTCGAGNAGAAACCCGAAAGCGGAGAAACAGAGTCTCCAAACCAACAGAGAAC 2228
 QY 2131 AGAGAAACACCTGTGAGCCGAGAGTCCCTCAAGTAGAGACTGAAAGAAAGTAGAAGCCCA 2190
 Db 2229 AGAAGAATCACCAGAGGAATCAGAGAAACCTCAGGTCGAGACTGAAAGAGTTGAGAGAAA 2288
 QY 2191 ACTCAAGAGACAGAGTTTCTGCTGCGAAGAGTAAGCGGATCTAGTCTGAAAGCCCAATGC 2250
 Db 2289 ACTGAGAGGCTGAAGATTTACTTTGGAAGAAATCCAGGATCCAATTTATCAAGTCCCAATGC 2348
 QY 2251 AACAGAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAAATTTATGATAACCAATAG 2310
 Db 2349 CAAGAGACTCTCAGAGGATTAAGAAATAATTTACTATTGTCACCCAGAGCAACATATAC 2408
 QY 2311 TATCATGGCAGAGCAGAGAAATTTACTTGGCTGTTGTTAAAGGAGTAA 2358
 Db 2409 TATTATGGCAGAGCTGAAAAACTATTGGCTTTATTAAAGGAGAGTAA 2456

RESULT 15

AAA47605
 ID AAA47605 standard; DNA; 2531 BP.

AC AAA47605;

DT 20-OCT-2000 (first entry)

DE Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.

OS Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.

OS Streptococcus pneumoniae.

PH Key Location/Qualifiers
 FT CDS 1..2531
 FT /*tag= a
 FT /product= "Sp36B polypeptide"

PN WO200037105-A2.

XX 29-JUN-2000.

PF 21-DEC-1999; 99WO-US030390.

PR 21-DEC-1998; 98US-0113048P.

PA (MEDI-) MEDIMUNE INC.

PI Johnson LS, Koenig S, Adamou JE;

XX WPI; 2000-452129/39.

DR P-PSDB; AAB01469.

XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.

XX Disclosure; Page 69-70; 70pp; English.

XX Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of S. pneumoniae in order to utilize such polypeptides in vaccines to

CC protect against a wide variety of S. pneumoniae. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHx) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections

XX Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 0 U; 3 Other;

Query Match 41.5%; Score 991; DB 3; Length 2531;
 Best Local Similarity 65.7%; Pred. No. 6.9e-240;
 Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGCACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTGTTC 57
 Db 60 TTCTTACGAGTTGGCACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTGTTC 119
 QY 58 CTATATAGATGGAACCAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGTTAG 117
 Db 120 TTATATAGATGCTGATCAGGCTGGTCAAAGGCGAGAAACTTTGACACAGATGAAGTCAG 179
 QY 118 CAAGCGTGAAGCAATCAATGCTGAGCAATCTCATCAAGATAACAGACCAAGGCTATGT 177
 Db 180 TAAGAGGGAGGGGATCAAGCGCGAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT 239
 QY 178 CACTTTCACATGGCGACCACTATCAATTTATCAATGTAAGTTCCTTATGAGCTATCAT 237
 Db 240 GACCTCTCATGGAGACCATTAATCAATTAATGCGAAGGTTCCCTATGATGCCATCAT 299
 QY 238 CAGTGAAGATTTACTCATGAAGATCCAAACTATAAGCTTAAAGATAGAGATATTGTTAA 297
 Db 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTGTCAA 359
 QY 298 TGAGTCAAGGCTGATGTTATCAAGTAGATGGAATTAATTAATTAATTAATTAATTAATTA 357
 Db 360 TGAATTAAGGCTGATGTTATCAATTAAGTAAACCGGTAATAATTAATTAATTAATTAATTA 419
 QY 358 TGCTGCCACGCGGATAAGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
 Db 420 TGCRCCTCATGGGATATATTTGCGACAAAGAGAGATTAACCTCAGACAGCAAGC 479
 QY 418 TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGCTGTGCTTGGCCAGCTTC 477
 Db 480 CAGTCAATATCAT-----AACTCAAGACAGATAATGCTGTGTCGACCCAGAGC 530
 QY 478 GCAAGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGA 537
 Db 531 CCAGGAGCTTATACACGGATGATGGGTATATCTTCAATGATCTGATATCAITGAGGA 590
 QY 538 TACTGCTGATGCTTATATCTGTTCCCTCATGGAGATCAATACCAATTACATTCCTAAGAATGA 597
 Db 591 CACGGGTGATGCTTATATCTGTTCCCTCAGCGGACCATTAACCATTAATTCCTAAGATGA 650
 QY 598 GTTATCAGTACGAGTTGGCTGCTGAGAGCTTCCTTATCTGTCGAGGAAATCTGTTC 657
 Db 651 GTTATCAGTACGAGTTAGCTGCTGAGAGCTTATTGGAATGG----- 695
 QY 658 AAATTCAGAGAACCTTATCCCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACC 717
 Db 696 -GAAGCAGGGATCTGCTCTTCTTCAAGTTCTAGTTTATAATGCAATCCAGCTCAACCAA 754
 QY 718 TTCTTAAGCAATCCAGGAACTACAAATACTAAACAAGCAACACAGCAACACTTAACAG 777
 Db 755 GATTGTCAGAGNACCAATCT-----GACTGCTACTCCAACTTA 794
 QY 778 TCAAGCAAGTCAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
 Db 795 TCATCAAAATCAAGGGGAGAAACATTTCAAGGCTTTTACGTGAATTTGATGCTTAACCCCTT 854
 QY 838 GAGTCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCACAATAACAGAGTCG 897

Db 855 ATCAGAACCCATGTGGAATCTGATGCGCTTATTTTCGACCCAGCGCAATCAAGTCTG 914
QY 898 AACAGCTAGAGGTGGTTCAGTGGCCACAGGAGATCATTAACACTTATCCCTTACTCTCA 957
Db 915 AACCGCAGAGGTGTAGTCTGCTCATGTGTAAACCATTAACCACTTTATCCCTTATGAACA 974
QY 958 AATGTCTGAATTTGGAAGAACGATCGCTGATTAATTTCCCTTCGTTATCGTTATCGTTCAACCA 1017
Db 975 AATGTCTGAATTTGGAAGAACGATTTGCTCGTATTAATTTCCCTTCGTTATCGTTCAACCA 1034
QY 1018 TTGGGTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCACACCGACTCCGGAAACCTAGTCC 1077
Db 1035 TTGGGTACCAAGATTCAAGCAGAGAACCAACCAAGTCCACACCGACTCCAGAACCTAGTCC 1094
QY 1078 AGGCCGGAACCTGCAACCAATCTTAAATTAAGATCAAAATCTTCTTTGGTTAGTCACT 1137
Db 1095 AAGTCCGCAACAGCTCCAGCAATCAATTTGATGGAA-----ATTGGTCAAGAGC 1148
QY 1138 GGTACGAAAGTGTGGGAAGGATATGTAATTCGAAGAAAGGGATCTCTCGTTATGCTT 1197
Db 1149 TGTTGAAAGATGAGCGCATGTTAIGCTTTTGAGAGATGGAGTTCTCGTTATATCCC 1208
QY 1198 TGCAGAAAGATTACCATCTGAACTGTAAATCTTTAAAGCAAGTTATCAAAACAAGA 1257
Db 1209 AGCCAAGATCTTTCAGCAGAAACAGCAGCAGGCAATGTAGCAAACTGGCCAGCAGGA 1268
QY 1258 GAGTGTTCACACACTTTAACTGTCTAAAGAGAAATGTTGCTCTCGTGACCAAGATT 1317
Db 1269 AAGTTTATCTCATAGCTAGGAACTAAGAAACTGACCTCCCATCTAGTATCAGAAAT 1328
QY 1318 TTATGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAAGGG 1377
Db 1329 TTACAATAAGCCTTATGACCTTACGAGCAAGTTTACCAAGATTTACTTGTATATAAGG 1388
QY 1378 TCGTAATCTGATTTCCAAGCCTTAGACAAATTAATTAGAAGCTTTGAATGATGAATCGAC 1437
Db 1389 TCGACAAGTTGATTTTGGAGCTTTGGATAACCTCTGTGAACGACTCAAGGATGCTCAAG 1448
QY 1438 TAATAAGAAATTTGGTAGATGATTTATGGCATTCCTAGCACCANTTACCATCCAGA 1497
Db 1449 TGATAAAGTCAAGTTAGTGAAGATATCTTTGCTCTTCTAGCTCCGATTCGTCATCCAGA 1508
QY 1498 GCGACTTTGGCAACCAAACTCTCAAAATGAGTACTAGACAGCAAGTTCTGTTATGCTCA 1557
Db 1509 ACGTTTAGAATAACCAATGCGCAATTTACCTACCTGATGATGAGATCAAGTAGCCAA 1568
QY 1558 ATTAGCTGATAAGTATACAACTGAGATGGTTACATTTTTCATGAACATGATATAATCAG 1617
Db 1569 GTTGGCAGGCAAGTACACAGCAGAAAGCGGTTATATCTTTGATCCTCGTGATATAACCA 1628
QY 1618 TGATGAAGGAGATGCATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAAGA 1677
Db 1629 TGATGAGGGGATGCTATGTAATCTCAATATGACCCATAGCCACTGGATTAAAGAAAGA 1688
QY 1678 TAGCCTTTCTGATAAGGAAAAAGTTGCAAGCTCAAGCTTATTAAGAAAAAGGTATCCT 1737
Db 1689 TAGTTTCTGAGCTGAGAGCGGAGCCAGCCAGCTTATGCTTAAAGAGAAAGTTTAC 1748
QY 1738 ACCTCCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGCAGCTAT 1797
Db 1749 CCCTCCCTCGACAGACCATCAGGATTCAGGAAATCTGAGGCAAAAGAGCAGAGCTAT 1808
QY 1798 TTACAATCGTGAAGAGGAAAAACGAATTCCTCGTTCGATCCATATATGTTGA 1857
Db 1809 CTACAACGAGTGAAGAGCAGTGAAGAGGTCCTTGTATGCTTCAATCTTCA 1868
QY 1858 GCATACAGTTGAGGTTAAAAAGCGTAAATTTGATTTATTCCTCATAGGATCATTAACATAA 1917
Db 1869 ATATACTGTAAGAGTCAAAAGCGGTAGTTTAAATCATACCTCATTTATGACCATTAACATAA 1928
QY 1918 TATTAAATTTCTGGTTTGTATGATCACATACAAAGCTCAATGCTATACCTTGA 1977

Db 1929 CATCAAAATTTAGTGGTTTTCAGCAAGGCTTTATAGGCACTTAAGGGGTATATCTTTGA 1988
QY 1978 AGATTTCTTTTGGCAGCATTAAGTACTACGTAGAACCCCTGACGAACGTCCACATTTCTAA 2037
Db 1989 GGATCTTTTGGCCAGCTGTCAAGTACTATGTGCAACATCCAAAGAACGTCCGATTCAGA 2048
QY 2038 TGATGATGGGCAATCCCACTGAGCATGTGTAGGCAAGAAAGACCACTAGTGAAGATCC 2097
Db 2049 TAATGGTTTGGTTAAACGTAGCCAGCATGTTCAAAGAAACAAATAATGTCAGCTGATAC 2108
QY 2098 AATAAGAACTTCAAAGCGGATGA-----AGAGCCAGT 2121
Db 2109 CAATCAACGGAATAACCAAGCGGAGGAGAAACCTCAGACAGAAAAAACCCTGAGGAAGAAC 2168
QY 2122 -----AGAGCCAGT 2130
Db 2169 CCTCGAAGAGAAACCCGAAAGCGAGAAACAGAGTCTCCAAAACCAACAGGGAACC 2228
QY 2131 AGAGGAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAAGGTAGAGGCCA 2190
Db 2229 AGAAGATCAACAGAGGAATCAGAAGAAACCTCAGTCTGAGACTGAAAAAGGTTGAAGAAA 2288
QY 2191 ACTCAAGAGCAGAGATTTTGGTTCGAAAGTAACGATTTCTAGTCTGAAAGCCAATGC 2250
Db 2289 ACTGAGAGGCTGAGATTTACTTTGAAAAATCCAGGATCCAAATTTCAAGTCCAAATGC 2348
QY 2251 AACAGAACTCTAGCTGTTTACGAAATAATTTGACTCTTCAAATTTATGATACAAATAG 2310
Db 2349 CAAGAGACTCTCAGAGATTTAAAAATAATTTACTATTTGGCACCAGCAGCAACATAC 2408
QY 2311 TATCATGGCAGAGCAGAAAAATTTACTTGGCTTTTAAAAAGGAAGTAA 2358
Db 2409 TATTATGGCAGAGCTCAAAAACTATTGGCTTTTATTAAAGAGAGATAA 2456

RESULT 16

ABX06705
ID ABX06705 standard; DNA; 2517 BP.
XX AC ABX06705;
XX AC
DT 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX S. pneumoniae type 4 strain coding region #993.
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
XX ear infection; antinflammatory; antibacterial; immunostimulant;
XX auditory; respiratory; gene therapy; vaccine.
XX Streptococcus pneumoniae; type 4 strain.
XX OS
XX PN W020027021-A2.
XX PD 03-OCT-2002.
XX XX 27-MAR-2002; 2002WO-IB002163.
XX XX 27-MAR-2001; 2001GB-00007658.
XX (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Masighani V, Tettelin H, Fraser C;
XX WPI; 2003-040579/03.
XX P-PSDB; AEU01418.

New proteins and nucleic acid molecules from Streptococcus pneumoniae,
useful as medicaments for treating or preventing a disease or infection
due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
ear infection.

PS Claim 6; SEQ ID NO 1985; 56pp; English.

XX The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, CC treating a patient by administering the protein, DNA or antibody (in a CC composition), a kit comprising first and second primers, which are the CC nucleic acid cited above or fragments between nucleotides 8-100 of a CC sequence not defined in the specification, for amplifying a target CC the first primer is substantially complementary to the target sequence CC and the second primer is substantially complementary to the complement of CC the target sequence, and where the parts of the primers having CC substantial complementarity define the termini of the target sequence to CC be amplified, assay comprising contacting a test compound with the CC protein, and determining whether the test compound binds to the protein CC and a Streptococcus pneumoniae bacterium, where one or more genes CC encoding the proteins has been rendered inactive. The proteins, nucleic CC acid molecules, antibody and compositions are useful as medicaments for CC treating or preventing a disease or infection due to streptococcus CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis CC media or ear infection. They are also useful in developing vaccines, CC diagnostics and antibiotics. The methods are useful for identifying CC immunodominant proteins. The present sequence is one of the 2489 CC identified coding region from the genomic sequence. Note: The sequence CC data for this patent did not form part of the printed specification, but CC was obtained in electronic format directly from WIPO at CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to CC standardise OS field)

XX Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 U; 0 Other;

SQ Query Match 41.4%; Score 990; DB 7; Length 2517;
 Best Local Similarity 68.1%; Pred. No. 1.2e-239;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 1 TTCTTACGAGTGGGACTGTATCAGCTAGAGACGGTTAAGGAAA---TAATCGGTTC 57
 DB 60 TTCTTATGAACCTTGGTCGTCCACAGCTGGTCAAGGTTAAGAGAGTCTAATCGAGTTTC 119

QY 58 CTATATAGATGAAACAAAGCAGCAGCAAAACGGAGATTTGACTCCGTGAGGTTAG 117
 DB 120 TTATATAGTGTATCAGCTGGTCAAAGGAGCAAAACCTTGACACCGATGAAGTCAG 179

QY 118 CAAGCGTGAAGGAATCAATGTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
 DB 180 TAAAGAGGAGGGGATCAAGCCGGAACAAATCGTCATCAAGATTAAGGATCAAGTTATGT 239

QY 178 CACTTCACATGCGGACCACTATCATATTAACATGTTAAGGTTCCCTATGAGCTATCAT 237
 DB 240 GACCTCTCATGGAGACCAATATCATCTACTATATAAGCGAGGTCCTTATGATGCCATCAT 299

QY 238 CAGTGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAGATGAGGATATGTTAA 297
 DB 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAAATTAATCAGTTGAAGGATTCAGACATTTGTTCAA 359

QY 298 TGAGTGAAGGTTGATATGTTTCAAGGTAGATGGAATAACTATGTTTACCTTAAGGA 357
 DB 360 TGAATCAAGGTTGTTATGTTTCAAGGTAGATGGAATAACTATGTTTACCTTAAGGA 419

QY 358 TGCTGCCACGCGGATAAGCTCCGTACAAAGAGGAAATCAATTCGACAAAAACAGAGCA 417
 DB 420 TCGAGCTCATGCGGATAATATTTCCGACAAAGAGAGATTAAACGTCAGAAGCAGGAACA 479

QY 418 TAGTCAACATCTGAAGTGGAACTCCAGAAACGATGGTGTGCTTGGCAGCTTC 477
 DB 480 CAGTCAATATACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGCCAGAGC 533

QY 478 GCAAGGACGCTATACATGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGA 537

DB 534 CCAAGCACGCTATACAAACGATGATGGTTATATCTTCAATGCAATCTGATATCATTTGAGGA 593

QY 538 TACTGTGATGCTTATATCGTTTCTCTCATGGAGATCAATTACCAATTACATTTCTTAAGAAATGA 597

DB 594 CACGGGTGATGCTTATATCGTTTCTCTCATGGGACCAATTACCAATTACATTTCTTAAGAAATGA 653

QY 598 GTTATCAGCTAGCGAGTGGCTGCTGCGAGAAGCCTTCTCTATCTGCTCGAGGAAATCTGTC 657

DB 654 GTTATCAGCTAGCGAGTGGCTGCTGCGAGAAGCCT----- 688

QY 658 AAATTCAGAAACCTATCGCGGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACC 717

DB 689 -----ATTGGAATGGGAAGCAGGATCTCGTCTTCTTCAAGTTCTAGTTTATATGTC 740

QY 718 TTCTGTGAAGCAATCCAGGAACTACAAATACTAACACAAAGCAACAGCAACCACTAACAG 777

DB 741 AAATCCAGCTCAACCAAGATTGTGAGAAACCACAATCTGACTGTCACTCCAACCTATCA 800

QY 778 TCAAGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAACTGCTTT 837

DB 801 TCA---AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTTAAACCTTT 857

QY 838 GAGTCAACGACATGTAGAACTCTGATGGCCTTCTTTGATCCAGACAAATCACAAGTCG 897

DB 858 ATCAGAACCCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAATCACAAGTCG 917

QY 898 AACAGTAGAGTGTGTCAGTCCACACGAGATCATTAACCACTTCACTTCACTTCTCA 957

DB 918 AACCCGACAGGTGTAGTGTCTCATGGTAACCAATTAACCACTTATCCCTTATGAACA 977

QY 958 AATGCTGAATTTGGAAGACGATCGCTGATATTTCCCTTCTGTTATCGTTTCAACCA 1017

DB 978 AATGCTGAATTTGGAAGAACGAAATGCTGATATTTCCCTTCTGTTATCGTTTCAACCA 1037

QY 1018 TTGGGTACCAAGTTCAGGCGCAGAAACAACTAGTCCACAAACGACTCCGGAACCTAGTCC 1077

DB 1038 TTGGGTACCAAGTTCAGGCGCAGAAACAACTAGTCCACAACTGACTCCGGAACCTAGTCC 1097

QY 1078 AGGCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTC-----TTT 1125

DB 1098 AAGTCGCAACCTGCAACCAATCTTCACACGCTCCAGCAATCCAAATGTAGAGAAAT 1157

QY 1126 GGTAGTACGCTGCTAGCAAGATTTGGGAGGAGATATGTTTCGAGAAAGGATCTC 1185

DB 1158 GGTCAAGAGCTGTTTCGAAAAGTAGGCGATGTTATGTTTGGAGGAAATGGAGTTTC 1217

QY 1186 TCGTTATGCTTTTGGAAAAGATTTACCATCTGAAAACCTGTTAAATCTTGAAGCAAGTT 1245

DB 1218 TCGTTATATCCAGCCCAAGGATCTTTACGAGAAACAGCAGCAGGCAATGTAGCAACT 1277

QY 1246 ATCAAAAACAGAGAGTGTTCACACATTTTAACTCTGCTAAAAGAAAATGTTGCTCTCG 1305

DB 1278 GGCCAAAGCAGGAAAGTTTATCTCTATAAGCTAGGAGCTAAGAAAACCTGACCTCCACTAG 1337

QY 1306 TGACCAAGAAATTTTATGATAAGCAATATCTGTTTAACTGAGGCTCATTAAGCCTTGT 1365

DB 1338 TGATCGAATTTTACAATAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACT 1397

QY 1366 TGNAAATGAAGGTCGTAATTTCTGATTTCCAAAGCCTTTAGACAAATTTATTAGAACGTTGAA 1425

DB 1398 TGATAATAAGGTCGACAAAGTTGATTTTGGAGCTTTTGGATTAACCTGTTGGAACTGCTCA 1457

QY 1426 TGATGATCGACTAATAAAGAAAATTTGGTAGATGATTTATTTGGCAATTTCTTAGCAACCAAT 1485

DB 1458 GGATGCCCAAGTGTATAAAGTCAAGTTAGTGGATGATATTTCTTGGCTTCTTAGCTCCGAT 1517

QY 1486 TACCCATCCAGGCGACTTGGCAACCAAAATCTCAAAATGAGTATGATGATGAGAGAGT 1545

DB 1518 TCGTCATCCAGAACGTTTAGGAAAACCAAAATGCGCAAAATTAACCTACACTGATGATGAT 1577

QY 1546 TCGTATTCTCAATTAGCTGATAGTATACAAACGTCAGATGGTTACATTTTGTATGAACA 1605

DB 1578 TCAAGTACCCAAAGTTGGCAGGCAAGTACAAACAGAACGGTTATATCTTTGATCTCTCG 1637

QY	1606	TGATATAATCAGTGTGATGAAGAGATGCATATGTAAACGCCTCATATGTGGCCATATAGTCACTG	1666
Db	1638	TGATATAAACAGTGTATGAGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTG	1697
QY	1666	GATTGGAAAGATAGCCCTTTCTGATAAGGAAAGTTGACAGCTCAAGCCTATATAAAGA	1725
Db	1698	GATTAAAAAAGATAGTTGTCTGAAGCTGAGAGAGCGCAGCCAGCGCTTATGCTAAAGA	1757
QY	1726	AAAAGGTATCCTACCTCCATCTCCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAG	1785
Db	1758	GAAAGGTTTGACCCCTCCTTCGACAGACCATCAGGATTCAGGAAATACTCAGGCAAAAGG	1817
QY	1786	TGCAGCAGCTATTTTACAATCGTGTGAAAGGGGAAACGAATTCCACTCGTTCGACTTCC	1845
Db	1818	AGCAGAAGCTATCTTACAAACCGGTGAAAGCAGCTAAGAAGGTGCCACTTGATCGTATGCC	1877
QY	1846	ATATATCGTTGAGCATACAGTTGAGGTTTAAAAACGGTAAATTTGATTATTCTTCATAAGGA	1905
Db	1878	TTACAATCTTCAATATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCTCATTATGA	1937
QY	1906	TCATTACCAATAATTAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGG	1965
Db	1938	CCATTACCAATCAATCAAAATTTGAGTGGTTTGACGAAGGCGCTTTATGAGGCACCTAAGGG	1997
QY	1966	CTATACCTTGGAGATTTGTTTGGCGAGATTAAAGTAACTAGTAAACACCTGACGAACG	2025
Db	1998	GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGAAACATCCAAACGACG	2057
QY	2026	TCCACATTCTAATGATGGATGGGCAATGCCAGTGAGCATGTGTAGGCAAGAAAGACCA	2085
Db	2058	TCCGCAATTCAGATAATGGTTTGGTACGCTAGCGACCATGTTTCGTAATAAATAAGGTAGA	2117
QY	2086	CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGGAGGAACACCTGC	2145
Db	2118	CCAAGCAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCACTCACCC	2177
QY	2146	TGAGCCAGAGAAG	2156
Db	2178	TGAATCTGATG	2188
RESULT 17	AAAA7602		
ID	AAAA7602	standard; DNA; 2531 BP.	
AC	AAA47602;		
XX	20-OCT-2000	(first entry)	
DT	Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.		
DE	Streptococcus pneumoniae; infection; vaccine; coiled coil region;		
XX	histidine triad residue; Sp36; antibody; otitis media;		
KW	nasopharyngeal infection; bronchial infection; bronchitis; sepsis;		
KW	meningitis; lobar pneumonia; ds.		
XX	Streptococcus pneumoniae.		
OS	Key	Location/Qualifiers	
PH	CDS	1..2517	
FT		/**tag= a	
FT		/product= "Sp36D polypeptide"	
XX	WO200037105-A2.		
PN	29-JUN-2000.		
PD	21-DEC-1999;	99WO-US030390.	
XX	21-DEC-1998;	98US-0113048P.	
XX	(MEDI-) MEDIMUNE INC.		

Johnson LS, Koenig S, Adamou JE;
 WPI: 2000-452129/39.
 P-PSDB; AAB01466.
 Vaccine useful for prophylaxis and treatment of pneumococcal infections
 such as otitis media, nasopharyngeal and bronchial infections, comprises
 Streptococcus pneumoniae proteins.
 Disclosure; Page 57-58; 70pp; English.
 Although a number of proteins have been suggested as being involved in
 the pathogenicity of Streptococcus pneumoniae, there still remains a need
 to identify polypeptides having epitopes in common from various strains
 of S. pneumoniae in order to utilize such polypeptides in vaccines to
 protect against a wide variety of S. pneumoniae. New vaccine compositions
 are described which comprise a Streptococcus pneumoniae polypeptide (or
 fragments) of 80 - 680 amino acids in length that comprise at least one
 histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
 directed against these features. The vaccine is useful in protecting
 against infection by Streptococcus pneumoniae. The vaccine composition
 comprising antibodies to is useful for passive immunization for treating
 Pneumococcal infections which includes otitis media, nasopharyngeal and
 bronchial infections
 Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 U; 0 Other;

Query Match	41.4%;	Score	990;	DB 3;	Length	2531;
Best Local Similarity	68.1%;	Pred. No.	1.2e-239;			
Matches 1478;	Conservative	0;	Mismatches	636;	Indels	57; Gaps
QY	1	TTCTTACGAGTTCGGGACTGTATCAAGCTGACAAACGGTTAAGGAAAA	---	TAATCGTGTTC	57	
DB	60	TTCTATGACTTGGTCGTACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC		119		
QY	58	CTATATAGATGGAAAAACAACGACGACGAAAAACGGAGAAATTGACTCTCATGAGGTAG		117		
DB	120	TTATATAGATGGTGTATCAGGCTGGTCAAAAGGCGAGAAACTTGACACAGATGAAGTFCAG		179		
QY	118	CAACGGTGAAGGAATCAATCTCAGCAAAATCGTTCATCAAGATAACAGACCAAGGCTATGT		177		
DB	180	TAAAGGAGGAGATCAACGCCAACAATCGTCAICNAGATTACGGATCAAGGTTATGT		239		
QY	178	CACATTCATGGCGACCACTATCATTTATTAACAATPGTAAAGGTTCTTTATGACGCTATCAT		237		
DB	240	GACCTCTCATGGAGACCATTTATCTACTATAATGCAAGTCCCTTATGATGCCATCAT		299		
QY	238	CAGTGAAGATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGATATTGTTAA		297		
DB	300	CAGTGAAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGATTTCAGACATTTGCA		359		
QY	298	TGAGGTCAGGGTGGATGTCTTCAAGGTAGATGGAAAACTATGTTTACCTTAAGGA		357		
DB	360	TGAATCAAGGGTGGTATGTATCAAGGTAGATGGAAAACTATGTTTACCTTAAGGA		419		
QY	358	TGCTGCCACGCGGATACGTCGCGTCAAAAGAGGAAATCAATCGACAAAACAAGAGCA		417		
DB	420	TGCAGCTCATCGCGATTAATTCGGAACAAAGAGAGATTAAACGTGAGAGCAGGAACA		479		
QY	418	TAGTCAACATCGTGAAGGTGAACTCCAGAAAAAGATGGTGTGCTTGGCCCTTGGCACGCTTC		477		
DB	480	CAGTCATAATCAACGGGGTGGTCT-----ACGATCAAGCAGTAGTTGCAGCCAGAGC		533		
QY	478	GCAAGGACGCTATACATACAGATGAGTTTATCTTTTAATGCTTCTGATATCATAGAGGA		537		
DB	534	CCAAGGACGCTATACAAACGGATGATGTTTATCTTCAATGCATCTGATATCATTTAGGA		593		
QY	538	TACTGGTGTGCTTATATCGTTCTCATGAGATCATTTACCATTTACATTTCTTAAGAATGA		597		
DB	594	CACGGGTGATGCTTATATCGTTCTTCTACGGGACCAATTCATTTACATTTCTTAAGAATGA		653		
QY	598	GTTATCAGCTAGCGAGTTGGCTGCTGAGAGGCTCTCTATCTGGTGCAGCAAAATCTGTC		657		

```

Db      654  GTTATCAGCTAGCGAGTGTAGTCTGCTGAGAGCCT----- 688
QY      658  AATTCAGAACTATCGCGGCAAAATACGATATACATTCAGAAACAACTGGGTACC 717
Db      689  -----ATTGGAATGGGAAGCAGGATCTCGTCTCTTCTCAAGTTCTAGTTATAATGC 740
QY      718  TTCTGTAAGCAATCCAGGAACATAAATACTAACACAGCAACAAACAGCAACTAACAG 777
Db      741  AATTCAGCTCAACCAAGATTGTACAGAACCAATCTGACTGTCTCCTCAACTTATCA 800
QY      778  TCAAGCAAGTCAAAGTAATGACATTTGATATGATCTCTTTGAAACAGCTCTTCAAACTGCTTT 837
Db      801  TCA---AATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTTAAACCCCT 857
QY      838  GAGTCAACGACATGTAGATCTGATGGCCTTGTCTTTGATTCAGACCAAAATCAAGTCG 897
Db      858  ATCAGAACCCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAAATCAAGTCG 917
QY      898  AACAGCTAGAGGTGTTGCAAGTGCACACAGGAGATCAATACCACTTCCATCCTTACTCTCA 957
Db      918  AACCCGACAGGTGTAGCTGTCCCTCATGTTAACCAATTACCCTTTATCCCTTATGAACA 977
QY      958  AATGCTGTAATTTGGAAGAACGAATCGCTCGTATTTATTCCTTCCTTGTTCATCAAAACA 1017
Db      978  AATGCTGTAATTTGGAAGAACGAATTCGCTCGTATTTATTCCTTCCTTGTTCATCAAAACA 1037
QY      1018  TTGGGTACAGATTCAGGCGCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db      1038  TTGGGTACAGATTCAGGCGCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1097
QY      1078  AGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTC-----TTT 1125
Db      1098  AAGTCGCAACCTGACCAAAATCTCACAACGCTCCAGCAATCCAATGTAGAGAAAT 1157
QY      1126  GGTATGCTAGCTGTGACGAAAGTTGGGGAAGATATGATTGGAAGAAAGGCAATCTC 1185
Db      1158  GGTCAAAGAGCTGTTTCGAAAGTAGCGATGTTATGTTCTTTGAGGAGAAATGGAGTTTC 1217
QY      1186  TCGTTATGCTTTGCAAGAGATTACCATCTGAACTGTAAATCTTGAAGAGAAAT 1245
Db      1218  TCGTTATATCCGACCAAGATCTTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACT 1277
QY      1246  ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATTTGCTCCTCG 1305
Db      1278  GGCCAGCAGGAAAGTTTATCTCATAGCTAGAGCTAGAAAACCTGACTCCCATCTAG 1337
QY      1306  TGACCAAGAAATTTATGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGT 1365
Db      1338  TGATCGAGAATTTTACAATAAGCCTTATGACTTACTAGCAAGAATTCACCAAGATTTACT 1397
QY      1366  TGNAAATAGGCTCGTAAATCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAA 1425
Db      1398  TGATATAAAGTCGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAAACGACTCAA 1457
QY      1426  TGATGAATCGACTATAATAAGAAAAATTTGTAGATGATTTATTTGGCATCTAGCACCAT 1485
Db      1458  GGAATGTCCTCAAGTATAAAGTCAAGTTAGTGATGATATTTCTGCTTCTTAGCTCCGAT 1517
QY      1486  TACCATTCAGAGGACTTGGGAAACCAAAATCTCAAATTTAGTATGATTAAGAGAGAGT 1545
Db      1518  TCGTTCATCCAGAACGTTTAGGAAAAACCAATGGCAATTTACCTACACTGATGAGAT 1577
QY      1546  TCGTATTCCTAATAGCTGATAAGTATACAGCTCAGATGTTTATCTTTTATGATGAACA 1605
Db      1578  TCAAGTAGCAAGTTGGCAGGCAAGTATACACAGAAAGCGGTATATCTTTTATCTCTCG 1637
QY      1606  TGATATAATCAGTATGAAGGAGATGCATATGTAAGCCTCATATGGGCAATAGTCACTG 1665
Db      1638  TGATATAACAGTATGAGGGGATGCTCTATGTTACTCCACATATGACCCATAGCCACTG 1697
QY      1666  GATTGGAAGATAGCCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATACTATAAGA 1725

```

```

Db      1698  GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATCTAAAGA 1757
QY      1726  AAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAATCTGAGATAG 1785
Db      1758  GAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAG 1817
QY      1786  TGACGACGCTATTTCACATCGTGTGAAGGGGAAAAACGAATTCCTACTGTTCCGACTTC 1845
Db      1818  AGCAGAAAGCTATCTCAACCGGTGAAGAGCGCTAAGAAAGGTGCCACTTGTATCGTATGC 1877
QY      1846  ATATATGTTGAGCATACATAGTTGAGGTTAAAAACGGTAATTTTATTTATCTCTCATAGGA 1905
Db      1878  TTACAATCTTCAATATCTAGTAGAGTCAAAACGGTAGTTTAACTCATCTCATTTATGA 1937
QY      1906  TCATTAACCAATATTTAAATTTGCTTGTGTTGATGATCAACATACAAAGCTCCAAATGG 1965
Db      1938  CCATTAACCAATACATCAAAATTTGAGTGGTTTGACGAAGGCCCTTTATGAGGCACCTAAGG 1997
QY      1966  CTATACCTTTGGAAGATTTGTTTGCAGAGATTAAGTACTAGTAGACACCTGACCGAAG 2025
Db      1998  GTATATCTTGTAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAGC 2057
QY      2026  TCCACATCTTAATGATGATGGGCAATGCCATGAGCATGTTTAGGCAAGAAAGACCA 2085
Db      2058  TCGCATTCAGATAATGTTTGGTAAACCTAGCGACCATGTTTCGTAATAAATAGGTAGA 2117
QY      2086  CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGACCGTAGAGGAAACACCTGC 2145
Db      2118  CCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAAGTGAGCAACTCACCC 2177
QY      2146  TGAGCCAGAAG 2156
Db      2178  TGAATCTGATG 2188

```

RESULT 18

ABS56454.08
Continuation (9 of 22) of ABS56454 from base 800001 (Streptococcus pneumoniae type 4 str)
WP Sequence split into 22 fragments LOCUS ABS56454 Accession Abs56454

WP	Fragment Name	Begin	End
WP	ABS56454_00	1	110000
WP	ABS56454_01	100001	210000
WP	ABS56454_02	200001	310000
WP	ABS56454_03	300001	410000
WP	ABS56454_04	400001	510000
WP	ABS56454_05	500001	610000
WP	ABS56454_06	600001	710000
WP	ABS56454_07	700001	810000
WP	ABS56454_08	800001	910000
WP	ABS56454_09	900001	1010000
WP	ABS56454_10	1000001	1110000
WP	ABS56454_11	1100001	1210000
WP	ABS56454_12	1200001	1310000
WP	ABS56454_13	1300001	1410000
WP	ABS56454_14	1400001	1510000
WP	ABS56454_15	1500001	1610000
WP	ABS56454_16	1600001	1710000
WP	ABS56454_17	1700001	1810000
WP	ABS56454_18	1800001	1910000
WP	ABS56454_19	1900001	2010000
WP	ABS56454_20	2000001	2110000
WP	ABS56454_21	2100001	2162598

Query Match 41.4%; Score 990; DB 7; Length 110000;
Best Local Similarity 68.1%; Pred. No. 6.3e-239;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

```

QY      1  TTCTTACAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAA---TAATCGTGTTC 57
Db      38008  TTCTTATGAACCTTGGTCTGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 38067
QY      58  CTATATAGATGGAACCAAGCGACCAAAAAACCGAGAAATTTGACTTCCTGTAGAGTTAG 117

```

Db	38068	TTATATAGATGGTGATCAGGCTGGTCAAAAGCGCAAAACTTGTACACAGATGAAGTCAAG	38124
Qy	118	CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGATAACAGACCAAGCTATGT	177
Db	38128	TAAAGGGAGGGGATCAACGCCGAACAATCGTCATCAAGATTACGGATCAAGGTTATGT	38187
Qy	178	CAC TTCACATGGCGGACACTATCATTTATTA CAATGGTAAGGTTCTTATGACGCTATCAT	237
Db	38188	GACCTCTCATGAGACCATTTATCATTTACTATTAATGCAAGGTCCCTTATGATGCCATCAT	38247
Qy	238	CAGTGAAGAATTACTCATGAAAGATCCAAACTATTAAGCTAAAAGATGAGGATATTGTTAA	297
Db	38248	CAGTGAAGACTCTCTCATGAAAGATCCGAATTTATCAGTTGAAGNATCAGACATTGTCAA	38307
Qy	298	TGAGGTCAAAGGTGGGATATGTTATCAAGGTAGATGGAAATACTATGTTTACTTCAAGGA	357
Db	38308	TGAAATCAAGGGTGGTTATGTTTATCAAGGTAGATGGAAATACTATGTTTACTTCAAGGA	38367
Qy	358	TGCTGCCACCGCGGATAACGTCGTCACAAAGAGAGAAATCAATCACAACAAAACAAGAGCA	417
Db	38368	TGCAGCTCATCGCGGATAAATTCGGAACAAAGAGAGATTAACCTCAGACGAGGAAACA	38427
Qy	418	TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTGCTTGGCACGTTTC	477
Db	38428	CAGTCATAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTTGCAGCCAGAGC	38481
Qy	478	GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGGA	537
Db	38482	CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGATCTGATATCATTTAGGA	38541
Qy	538	TACTGGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATTTACATTTCTTAAGAATGA	597
Db	38542	CACGGTGATGCTTATATCGTTCCTCAGCGGACCATTTACATTTACATTTCTTAAGAATGA	38601
Qy	598	GTTATACGCTAGCGAGTTGGCTGCTCGAGAGCCTTCCCTATCTGGTCGAGGAATCTGTC	657
Db	38602	GTTATACGCTAGCGAGTTAGCTGCTCGAAGCCT-----38636	
Qy	658	AAATTCAGAAACCTATCGCGGACAAAATAGCGATACACTTCAAGAAACAAACTGGGTACC	717
Db	38637	-----ATTGGAATGGGAAGCAGGAATCTGTCCTCTTCAAGTTCTAGTTATATATGC	38688
Qy	718	TTCTGTAAAGCAATCCAGGAACATAAATACTAACACAGCAACAACAGCAACACTTAACAG	777
Db	38689	AAATCCAGCTCAACCAAGATTGTGAGAAACCAATCTGACTGTCACTCCAACTTATCA	38748
Qy	778	TCAGCAAGTCAAGTAATGACATTTGATGCTCTTGAACAGCTCTCAAACTGCCCTTT	837
Db	38749	TCA---AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTT	38805
Qy	838	GAGTCAACGACATGTAGAATCTGATGCCCTTGTCTTGTATCCAGCAACAATCACAAGTCG	897
Db	38806	ATCAGAACGCCATGTGGAATCTGATGSCCTTTATTCGACCCAGCGCAATCACAAGTCG	38865
Qy	898	AACAGCTAGAGGTGTTCCAGTGCCACACGAGATCATTTACCACTTATCCCTTACTCTCA	957
Db	38866	AACGCCAGAGGTGTAGCTGTCCCTCATGTAACCAATACCCTTTATCCCTTATGAACA	38925
Qy	958	AATGCTGAATTGGAAGAACGAATCGCTCGTATTTATTCCTCCCTTCGTTATCGTTCAAAACA	1017
Db	38926	AATGCTGAATTGGAAGAACGAATTGCTCGTATTTATCCCTTCGTTATCGTTTCAAACA	38985
Qy	1018	TTGGGTACCAAGTTCAGGGCAGAAACAACCAAGTCCAAACCGACTCCGGAACCTAGTCC	1077
Db	38986	TTGGGTACCAAGTTCAGAGACAGAAACAACCAAGTCCCAATCGACTCCGGAACCTAGTCC	39045
Qy	1078	AGGCCCGCAACCTGCACCAATCTTAAAAATAGACTCAAAATTTTC-----TTT	1125
Db	39046	AAGTCCGCAACCTGCACCAATCTTCAACAGCTCCAGCAATCCAAATTTGATGAGAAATT	39105
Qy	1126	GGTTAGTCAGCTGGTACGAAAGTTGGGGAAGGATATGTATCGAAGAAAGGCGATCTC	1185
Db	39106	GGTCAAGAAGCTGTTGAAAGTAGGCCATGGTTATGTTCTTTGAGGAATAGGAGTTTC	39165

QY 1246 ATCAAAACACAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATCTTGCTCCTCG 1305
 Db 1219 GGCCACAGCAAGAAAGTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG 1278
 QY 1306 TGACCAAGAAATTTATGATTAAGCATATATCTTTAACTGAGGCTCATAAAGCTTTGT 1365
 Db 1279 TGATCGAATTTTACAAATAGGCTTATGACTTACTAGCAAGATTCACCAAGATTACT 1338
 QY 1366 TGNAAATPAGGGTCGTAAATCTGATTTCCAGCCTTAGACAAATATTAGACCTTGAA 1425
 Db 1339 TGATAATAAAGGTCGACAAAGTTGATTTTGAAGCTTTGGATAACCTGTTGGAACGACTCAA 1398
 QY 1426 TGATGATCGACTAATAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAAT 1485
 Db 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATCTTCCCTTCCTTAGCTCCGAT 1458
 QY 1486 TACCCATCCAGAGCGACTTTGGCAACCAAAATTTCTCAAATTTGAGTATACTGAAGACGAAGT 1545
 Db 1459 TCGTCATCCAGAACGTTTAGGANAACCAATGCGCAAAATTTACCTACACTGATGATGAT 1518
 QY 1546 TCGTATTTGCTCAATAGCTGATAGTATACAACTGAGATGTTTACATTTTGTGATGACA 1605
 Db 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACGAGACGGTTTATATCTTTGATCCTCG 1578
 QY 1606 TGATATAATCAGTCATGAAGAGATGATATGTAACGCTCATATGGGCCATGCTCACTG 1665
 Db 1579 TGATATAACCAAGTATGAGGGGATGCTTATGTAACCTCCATATGACCCATAGCCACTG 1638
 QY 1666 GATTGGAAAAGATAGCCTTTCTGATAGGAAAAAGTTGACGCTCAAGCCTATACATAAGA 1725
 Db 1639 GATTAAAAAGATAGTTGCTGAGCTGAGAGAGCGCAGCCAGGCTTATGCTAAAGA 1698
 QY 1726 AAAAGGTATCCTACTCCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAG 1785
 Db 1699 GAAAGGTTTGAACCTCCTTCCAGACAGACCATCAGGATTCAGGAAATACCTGAGGCAAAAGG 1758
 QY 1786 TGCAGCAGCTATTACATATCTGTAAGAGGGAAGAAACGAAATTCACCTGCTGCTGACTTCC 1845
 Db 1759 AGCAGAGCTATCTACAAACCCGTTGAAGAGAGCTTAAGAGGTGCGACTTGTGATGCTGCC 1818
 QY 1846 ATATATGTTGAGCATACAGTTGAGGTTTAAACCGTAAATTTGATTTCTCTCATAGA 1905
 Db 1819 TTACAATCTTCAATATACTGTAGAGTCAAAACCGTAGTTTAAATCATACCTCATTAAGA 1878
 QY 1906 TCATTACCAATTAATTAATTTGCTTGTGTTGATGATCAACATACAAAGCTCCAAATGG 1965
 Db 1879 CCATTACCAATTAATTAATTTGCTTGTGTTGATGATCAACATACAAAGCTCCAAATGG 1938
 QY 1966 CTATACCTTGAAGATTTGTTTGGACGATTAAGTACTAGTACGTAACACCTGACGAAACG 2025
 Db 1939 GTATACCTTGAAGATTTGTTTGGACGATTAAGTACTAGTACGTAACACCTGACGAAACG 1998
 QY 2026 TCCATTTCTAATGATGGATGGGCAATGCCAGTGAGCATGTTTGGGCAAGAAAGACCA 2085
 Db 1999 TCCGATTTCAAGATGTTTGGTGAACGCTAGCAGCCTGTTCAAGTACTATGCGNACATCAACGACG 2058
 QY 2086 CAGTGAAGATCAATAGAACTTCAAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCG 2145
 Db 2059 TCAAGCTGATACCAATCAACCGAAAAACCAAGGAGGAGAAACCTCAGACAGAAAAACC 2118
 QY 2146 TGAGCCGAAGTCCCTCAAGTAGAGA 2171
 Db 2119 TGAGGAAGAAACCTTCGAGAGAGA 2144

RESULT 20
 ID ABQ84824
 XX ABQ84824 standard; DNA; 2290 BP.
 AC ABQ84824;
 XX
 DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; gene; ds.
 XX Streptococcus pneumoniae.
 XX US2002061545-A1.
 XX 23-MAY-2002.
 XX 22-JAN-2001; 2001US-00765272.
 XX 30-OCT-1997; 97US-00961083.
 XX (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX Choi GH, Kunsch CA, Barash SC, Dillon PU, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2002-479261/51.
 DR P-PSDB; ABP54589.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX Claim 1; Page 28-29; 70pp; English.
 XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;
 SQ
 Query Match 41.3%; Score 987.6; DB 6; Length 2290;
 Best Local Similarity 67.7%; Pred. No. 4.8e-239;
 Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
 QY 1 TTCTTACGAGTTGGACCTGTATCAAGCTAGAACGGTTAAGGAAAA--TAATCGGTGTTTC 57
 Db 4 TTCTTATGAATCTGGTCTGTCACCAAGCTGGTTCAGGTAAAGAAAGAGTCTAATCGAGTTTC 63
 QY 58 CTATATATAGTATGAAAAAAGCGAGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTTAG 117
 Db 64 TTATATAGTATGATCAGGCTGGTCAAAAGGCGAGAAACTTGACCCAGATGAAGTCAG 123
 QY 118 CAAGCGTGAAGGATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGT 177
 Db 124 TAAGAGGGAGGGGATCAACGCCGAACAAATGTTATCAAGATTACGGATCAAGGTTATGT 183
 QY 178 CACTTCATCGCGACCACTATCATTTATCAATGGTAAAGTTCCTTATGACGCTATCAT 237
 Db 184 GACCTCTCATGAGACCAATTAATCAATTAATGCGAAGGTTCTTATGATGCCATCAT 243
 QY 238 CAGTGAAGAAATTAATCAATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
 Db 244 CAGTGAAGAGCTCTCTCATGAAAGATCCGAATTAATCAAGTTATCAAGTTGAAGGATTCAGACATTTGTC 303
 QY 298 TGAGTCAAGGTTGATATGTTATCAAGGTAGATGTTATCAAGGTAGATGTTTACCTTAAGGA 357

Db 304 TGAATCAAGGTGGTATGATGATTAAGGTAAACGGTAATACTATGTTACCTTAAGGA 363
 QY 358 TCGTCCCAAGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAGCAAGAGCA 417
 Db 364 TGCAGCTCATGCGGATAATATTCGGACAAAGAGAGATTAACGTCAGAAAGCAGGAAGC 423
 QY 418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAACGATGGTGGTGGCCCTGTCACGTTTC 477
 Db 424 CAGTCATATCAT-----AACTCAAGAGCAGATAATGCTGTTGCTGCGACCGAGAGC 474
 QY 478 GCAAGACGCTACTACTACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGGA 537
 Db 475 CCAAGACGCTTATACAAACGATGATGGTATATCTTCAATGCATCTGATATCATAGAGGA 534
 QY 538 TACTGTGATGCTTATATGTTCTCATGAGAGATCATTAACCAATTAATTCCTTAAGAAATGA 597
 Db 535 CACGGGTGATGCTTATATGTTCTCATGCGGACCATTAACCAATTAATTCCTTAAGAAATGA 594
 QY 598 GTTATCAGCTAGCGAGTTCGCTGCTCAGAAAGCCTTCTCTATCTGGTTCGAGGAATCTGTC 657
 Db 595 GTTATCAGCTAGCGAGTTCGCTGCTCAGAAAGCCT----- 629
 QY 658 AAATCAAGAACTTATCGCCGCAAAATAGCGATACACTTCAAGAAACAAACTGGGTACC 717
 Db 630 -----ATTGGAATGGGAAGCAGGATCTCGTCTCTTCTCAAGTTCCTAGTTATATGTC 681
 QY 718 TTCTGTAGCAATCCAGGAACCTACAAATACTAACAGCAACAGCAACAGCAACCTAACAG 777
 Db 682 AAATCCAGCTCAACCAAGATTTGTCAGAGAACCAACCAATCTGACTCTCAACCTATATCA 741
 QY 778 TCAAGCAAGTCAAGATTAATGACATGATAGTCTCTTGAACAGCTCTACAACTGCTTT 837
 Db 742 TCA-----AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTTGCTAAACCTTT 798
 QY 838 GAGTCAACGACATGTAGAACTGTATGGCCCTTGTCTTTGATCCAGACACAAATCACAAAGTCG 897
 Db 799 ATCAGAACCCATGTGGAATCTGATGGCTTATTTTCGACCAGCGCAATCACAAAGTCG 858
 QY 898 ACAGCTAGAGGTGTGAGTGGCCACAGGAGATCATTAACCACTTCACTCCCTTACTCTCA 957
 Db 859 AACCCGCGAGGTGTAGCTGTCCCTCATGTTAACCAATTAACCACTTTATCCCTTATGAACA 918
 QY 958 AATGCTGAATTTGGAAGAACGATCGCTGATATTTCCCTTCTGTTATCGTTTCAAAACCA 1017
 Db 919 AATGCTGAATTTGGAAGAACGATTTGCTGATATTTCCCTTCTGTTATCGTTTCAAAACCA 978
 QY 1018 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCACAAACGACTCCGGAACCTAGTCC 1077
 Db 979 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCACAAACGACTCCGGAACCTAGTCC 1038
 QY 1078 AGGCCGCGAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
 Db 1039 AAGTCCGCAACCTGCACCAAAATCTTCAACGAGCTCCAAAGAAATCCAAATTTGATGAGAAAT 1098
 QY 1126 GGTATGCTAGCTGGTACGAAAGTTGGGGAAGATATGATTTGGAAGAAAGGCGATCTC 1185
 Db 1099 GGTCAAGAGCTGTTGGAAGTAGGCGATGTTATGCTTTGAGGAGAAATGGAGTTTC 1158
 QY 1186 TCGTTATGCTTTTGGGAAAGATTTTACCAATCTGAAAATGTTTAAAGAAATTTGGAAGCAAGTT 1245
 Db 1159 TCGTTATATCCAGCCAGGATCTTTTCAGCAGAAACAGCAGCAGGCAATTTGATAGCAAACT 1218
 QY 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAATTTGTTCTCTCTCG 1305
 Db 1219 GCGCAAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAACTGACCTCCCACTAG 1278
 QY 1306 TGACCAAGAAATTTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGT 1365
 Db 1279 TGATCGAGAAATTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1338
 QY 1366 TGNAAATAAGGTCGTAATTTCTGATTTCCAAAGCCTTTAGACAAATTTATTAGAACGCTTGA 1425

Db 1339 TGAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATACCTGTTGGAACGACTCAA 1398
 QY 1426 TGATGAATCGACTATATAAAGAAAAATTTGGTAGATGATTTATTTGGGCAATTCCTAGCACAAT 1485
 Db 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATAATCTTGGCTTCTTAGTCCGAT 1458
 QY 1486 TACCCATCCAGAGCGACTTTGGCAAAACCAATTTCTCAAAATTTGAGTATACTGAAGACCAAGT 1545
 Db 1459 TCGTCATCCAGAACGTTTAGGAAACCAATGCGCAAAATTTACCTACACTGATGATGAT 1518
 QY 1546 TCGTATTTGCTCAATAGCTGATAAGTATACAACTGATGATGTTTATCATTTTGGATGAACA 1605
 Db 1519 TCAAGTAGCAGAGTTGGCAGGCAAGTACACACAGAAAGCGTTATATCTTTGATCCTCG 1578
 QY 1606 TGATATAATCAGTATGAAGAGATGTCATATGTAAGCCTCATATGCGGCATATAGTCACTG 1665
 Db 1579 TGATATAACAGTATGAGGGGATGCTTATGTAACCTCCACATATGACCCATAGCCACTG 1638
 QY 1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATATAAAGA 1725
 Db 1639 GATTAAGAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGCCAGGCTTTATGCTAAAGA 1698
 QY 1726 AAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAG 1785
 Db 1699 GAAAGTTTGACCCCTCTTTCGACAGACCATCAGGATTCAGGAATACTGAGGCAAAAG 1758
 QY 1786 TCGACGAGCTATTTACAAATCGTGTGAAAGGGAAAAAAGAAATTCACCTCGTTCGACTCC 1845
 Db 1759 AGCAAGAGCTATCTAACCCGCTGAAAGCAGCTAAGAAAGTGCACACTTGATCGTATGCC 1818
 QY 1846 ATATATGTTGAGCATACAGTTGAGGTTAAAAACGTAATTTGATTTATCTCTCATAGA 1905
 Db 1819 TTACAACTTTCAATATCTGTAGAAAGTCAAAACGCTAGTTTAATCATACCTCATATGA 1878
 QY 1906 TCATTACCAATAATTTAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGG 1965
 Db 1879 CCATTACCAATAACATCAAAATTTGAGTGGTTTTCAGGAAGCCTTTATGAGGCACTTAAGG 1938
 QY 1966 CTATACCTTGGAAAGATTTGTTGCGACGATTAAGTACTAGCTAGAAACACCTGACGAAAG 2025
 Db 1939 GTATATCTTGGAGTCTTTTGGCGACTGTCTCAAGTACTATGTGCAACATCTCAAAAGCAAG 1998
 QY 2026 TCCACATCTCTAATGATGGATGGGCAATGCCAGTGGAGATGTTTAGGCAAGAAAGACCA 2085
 Db 1999 TCCGATTCAGATAATGTTTGGTAACTGCTAGCGACCAATGTTCAAAGAAACAAATGG 2058
 QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGC 2145
 Db 2059 TCAAGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118
 QY 2146 TGAGCCAGAAAGTCCCTCAAGTAGAGA 2171
 Db 2119 TGAGGAAGAAACCCCTCGAGAAGAGA 2144

RESULT 21
 ADC45146
 ID ADC45146 standard; DNA; 2290 BP.
 XX
 AC ADC45146;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE S. pneumoniae DNA encoding antigen SP042.
 KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.
 XX Streptococcus pneumoniae.
 OS
 XX
 PN US6573082-B1.
 XX
 PD 03-JUN-2003.
 XX

PF 28-MAR-2000; 2000US-00536784.
 XX
 PR 31-OCT-1996; 96US-0029960P.
 PR 30-OCT-1997; 97US-00961083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,
 PI Rosen CA;
 XX WPI; 2003-764574/72.
 DR P-PSDB; ADC45147.
 XX
 PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by Streptococcus pneumoniae.
 XX
 PS Example 1; SEQ ID NO 65; 58pp; English.
 XX
 CC The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence encodes
 CC an S. pneumoniae antigenic protein.
 XX
 SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;
 Query Match 41.3%; Score 987.6; DB 9; Length 2290;
 Best Local Similarity 67.7%; Pred. No. 4.8e-239;
 Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
 Y 1 TTCTTACGAGTTGGAGCTGATATCAAGCTAGAACGGTTAAAGGAAAA---TAATCGTGTTC 57
 D 4 TTCTTATGAAGTTGGTGTGACCAAGCTGTGTCAGTTTGAAGAGAGTCTATTCAGATTTC 63
 Y 58 CTATATAGATGGAACCAAGCGACGCAACCAAGGAGAAATTTGACTCTGATGAGGTAG 117
 D 64 TTATATAGATGCTGATCAGGCTGCTGCAAGGCGAGAAAACTTGACACAGATGAATCAG 123
 Y 118 CAAGCGTGAAGGATCAATGCTGAGCAATCGTCAATCAAGTAACAGACCAAGCTATGT 177
 D 124 TAAAGGGAGGGGATCAACCGCAAAATNGTWTATCAAGATTACGGATCAAGGTTATGT 183
 Y 178 CACTTACATGGGACCACTATCATTTATCAATGGTAAAGTTCTTATGACGTATCAT 237
 D 184 GACCTCTCATGGAGACCATATCATTTACTATATGCGAAGTTCTTATGATGCCATCAT 243
 Y 238 CAGTGAAGATTACTCTAAGAAATCCAACTAATAGCTAAGGATGAGGATATTTGTTAA 297
 D 244 CAGTGAAGAGCTCTCATGAAGATCCGAATTTATCAGTTGAAGATTTCAGACATTTGCAA 303
 Y 298 TGAGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTACCTTAAGGA 357
 D 304 TGAATCAAGGGTGGTATGTTATGTTAAGGTAAACCGTAAATACTATGTTACCTTAAGGA 363
 Y 358 TGCTGCCACGCGATTAACGTCCTGACAAAGAGGAATCAATCGACAAAACCAAGAGCA 417
 D 364 TGAGCTCATGCGGATTAATTCGACAAAGAGAGATTTAAAGTCAAGAGCAGGAACG 423
 Y 418 TAGTCAACATGCTGAAGGTGAACTCCAAAGAAACGATGGTGTGCTGTGCTTGGCAGCTTC 477
 D 424 CAGTCATAATCAT-----AACTCAAGAGCAGATAAATGCTGTGCTGCGAGCCAGAGC 474
 Y 478 GCAAGGCGCTACTACGATGATGTTATCTTAAATGCTTCTGATATCATAGAGGA 537
 D 475 CCAAGGAGCTTATACAGGGATGAGGTATATCTTCAATGATCATGATATCATTTGAGGA 534

538 TACTGGTGATGCTTATATCGTTCTCATGGAGATCATTACATTACATTCTCTAAGAATGA 597
 535 CACGGGTGATGCTTATATCGTTCTCATGGGACCATTAACATTCATTCTTAAGAATGA 594
 598 GTTATCAGCTAGCAGGTGGCTGCTGCAGAAAGCTTCTCTATCTGGTGCAGAAATCTGTC 657
 595 GTTATCAGCTAGCAGGTAGCTGCTGCAGAAAGCT----- 629
 658 AATTCAGAAACCTATCGCGACAAATAGGATAACACATTCAGAACAAACTGGGTACC 717
 630 -----ATTGGAATGGGAAGCAGGGATCTGCTCTTCTCAAGTCTCTAGTTATATAATGC 681
 718 TTCTGTATGAACAAATCCAGAACTACAAATACTAACAAGCAACAAACAGCAACATAACAG 777
 682 AATCCAGCTCAACCAAGATTGTGCAGAGAACCAATCTGACTGTCTACTCCAATTTATCA 741
 778 TCAAGCAAGTCAAAAGTAAATGACATTTGATGTCTCTTGAACAGCTCTTACAAATCTGCTTT 837
 742 TCA---AATCAAGGGGAAAAACATTTCAAGCTTTTACGTGAATTTGATGCTAAACCTTT 798
 838 GAGTCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCAAAATCAAACTCG 897
 799 ATCAGAAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCAAACTCG 858
 898 AACAGCTAGAGGTGTTGCAGTGCACACGAGGATCATTAACACTTCTATCCCTTACTCTCA 957
 859 AACGCCACAGGTGAGTGTCTCTCATGTGTAACCATTAACACTTATCCCTTATGAACA 918
 958 AATGCTGAATGGAAGAAAGATTCGTGCTGATTTATTTCCCTTCTGCTTATGCTTCAAAACCA 1017
 919 AATGCTGAATGGAAGAAAGATTTGCTGCTTATTTATCCCTTCTGCTTATGCTTCAAAACCA 978
 1018 TTGGGTACACGATTTCAAGGCCAGAACCAACCAAGTCCACACGACTCGGAACTAGTCC 1077
 979 TTGGGTACACGATTTCAAGAACCAAGTCCACAACTCGACTCGGAACTAGTCC 1038
 1078 AGGCCCCCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTC-----TTT 1125
 1039 AAGTCCGCACTGCACCAAAATCTTCAACAGCTCCCAAGCAATCCAAATTTGATGAGAAAT 1098
 1126 GGTATGTCAGCTGATGAGAAAGTGGGAAAGATGATTTATCGAAGAAAGGCACTCTC 1185
 1099 GGTCAAAAGAGCTGTTGCAAAAGTAGGCTGATGTTATGCTTTGAGGAGAAATGAGGTTTC 1158
 1186 TCGTATGCTTTGCGAAAGATTTACCATCTGAACTGTTTAAATCTTCAAAACCAAGTT 1245
 1159 TCGTTATATCCAGCAGGATCTTTCGAGAAACAGCAGCAGGCAATGATAGCAAACT 1218
 1246 ATCAAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCG 1305
 1219 GGCCACAGCAGAAAGTTTATCTCATAGTAGGAGCTAAGAAACTGACCTCCCATCTAG 1278
 1306 TGACCAAGAAATTTATGATTAAGCAATATCTGTTAACTGAGGCTCATTAAGCCTTGT 1365
 1279 TGATCGAGAAATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTACT 1338
 1366 TGNAAATAAGGGTGAATTTCTGATTTTCCAGCCTTAGACAAATTTATAGAACGCTTGA 1425
 1339 TGATAATAAGGTCGACAAAGTTGATTTTGAAGCTTTGATTAACCTGTTGGAACCACTCAA 1398
 1426 TGATGAATCGACTAATAAGAAAGAAATTTGATGATGATTTATTTGGCATTTCTTAGCACAAT 1485
 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGAGATATTTCTTCCCTTCTTAGCTCCGAT 1458
 1486 TACCATCCAGAGCATTGGCAACCAAAATCTCAAAATGAGTATGATGATGAGT 1545
 1459 TCGTCATCCAGAGCTTTAGGAAACCAAAATGCGCAAAATACCTACACTGATGATGAT 1518
 1546 TCGTATTTGCTCAATTTAGCTGATAGTATACAACTCAGATGGTTTACATTTTGTATGAACA 1605
 1519 TCAAGTAGCCAGTTGCGAGCAGAGTACACACAGAGAGCGTTATATCTTTGATCTCTCG 1578


```
Db 911 TCA--AAATCAAGGGGAARACATTTCAAGCCTTTACGGTAATGTATGCTAAACCCCTT 967
QY 838 GAGTCAACGACATGTAGAAATCTGATGGCCCTTGCTTTGATGCCAGACAAATCAAGTCG 897
Db 968 ATCAGAACGCCATGTAGAAATCTGATGGCCCTTAATTTTCGACCCAGCGCAAAATCAAGTCG 1027
QY 898 AACAGCTAGAGGTGTTGTCAGTGCACACGAGAGATCAATACCACTTCATCCCTTACTCTCA 957
Db 1028 AACCGCCAGAGGTGTAGTCTCCCTCATGTATACCAATACCACTTTATCCCTTATGAACA 1087
QY 958 AATGTCTGAAATTTGGAAGAACGAATCGCTCGTATATTTATCCCTTTCGTTTATGTTCAAAACCA 1017
Db 1088 AATGTCTGAAATTTGGAAGAACGAATTCGTGTATTTATCCCTTTCGTTTATGTTCAAAACCA 1147
QY 1018 TTGGGTACAGATTCAGGCCAGACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db 1148 TTGGGTACAGATTCAGAACAGACAAACCAAGTCCACATCGACTCCGGAACCTAGTCC 1207
QY 1078 AGGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAATCTTC-----TTT 1125
Db 1208 AAGTCTGCAACCTGCACCAAAATCTTCAACAGCTCCAGCAATCCAATTCATGAGAAAT 1267
QY 1126 GGTGTAGTCACTGTCACGAAAGTTGGGGAAGGATATGTATTCGAGAAAGAGGCATCTC 1185
Db 1268 GGTCAAGAAAGCTGTTCCGAAAGTAGGCGATGGTTATGTCTTTGAGGAGAAATGGAGTTTC 1327
QY 1186 TCGTTATGCTCTTTCGGAAGATTTACCATCTGAAACGTGTTAAATCTTGAAGCAAGTT 1245
Db 1328 TCGTTATATCCAGCCCAAGATCTTTTCAGCAGAAACAGCAGCAGGCAATGATGACAACT 1387
QY 1246 ATCAAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAATGTTGCTCCTCG 1305
Db 1388 GGCCACAGCAGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCCATCTAG 1447
QY 1306 TGACCAAGAAATTTATGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
Db 1448 TGATCGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1507
QY 1366 TGNAANTAAGGTCGTATCTGATTTCCAGGCTTAGACAAATTTATAGAACCTTGAA 1425
Db 1508 TGATAATAAGGTCGACAAAGTTGATTTGAGGTTTGGATAACCTGTGTGAACACTCAA 1567
QY 1426 TGATGAATCCACTAATAAGAAAAATTCGTAGATGATTTATGGCAATTCCTAGCACCAAT 1485
Db 1568 GGATGCTCAAGTGATTAAGTCAAGTTAGTGTGATGATATCTTGCTTCTTAGCTCCGAT 1627
QY 1486 TACCCATCCAGAGCGACTTGGCAACCAAAATCTCAAAATTCAGTATGATGAGACGAAGT 1545
Db 1628 TCGTCATCCAGAACGTTTAGGAAAAACCAAAATGCGCAAAATACCTACACTGATGATGAGAT 1687
QY 1546 TCGTATTGCTCAATTAGCTGATAGTATACAACTCAGATGTTACATTTTGTATGAACA 1605
Db 1688 TCAAGTAGCAAGTTGGCAGCAAGTACACAAAGAAAGCGTTATATCTTTGATCTCTG 1747
QY 1606 TGATATAATCAGTGCATGAAGAGATGTCATATGTAAACGCTCATATGGGCCATAGTCACTG 1665
Db 1748 TGATATAACAGTGCATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTG 1807
QY 1666 GATTGGAAGATAGCCTTCTGATAGGAAAAAGTTGCAGCTCAAGCCTATATAAGA 1725
Db 1808 GATTAAAAAAGATAGTTGTCTGAAGCTGAGAGAGCGGACCCAGGCTTATGCTAAGA 1867
QY 1726 AAAAGGTATCTCACTCCATCTCAGACGCGAGATGTTAAAGCAATCCAACTGAGATAG 1785
Db 1868 GAAAGGTTTGACCCCTCTCTGACAGACCACCGAGATTCAGAAATCTAGGCAAAAG 1927
QY 1786 TGCAGAGCTATTTACAACTGTTGAAAGGGGAAAAAGAAATTCACCTGTTCCACTTCC 1845
Db 1928 AGCAGAAAGCTATCTACAAACCGGTGAAAGCAGCTAAGAAAGGTGCACTTATGCTATGCC 1987
QY 1846 ATATATGTTGAGCATACAGTTGAGGTTTAAAGCGTAATTTGATTTATCTCTCATAGA 1905
```

```
Db 1988 TTACAATCTTTCAATATATCTAGTAAGTCAAAAAACGGTAGTTTAAATCATATCTCATTAAGA 2047
QY 1906 TCATTACCAATAATTAATTTTCCTTGGTTTGATGATACACATACAAAGCTCCAAATGG 1965
Db 2048 CCATTACCAATAATTAATTTGAGTGGTTTGACGAAGGCTTTATGAGGCACTTAAGG 2107
QY 1966 CTATACCTTTGGAAGATTTGTTTTCGAGAGTAAAGTACTAGTAGAACACCCCTGACGAACG 2025
Db 2108 GTATAGTCTTGAGATCTTTTTCGGACTGTCAAGTACTATGTGCAACATCCAAACGAACG 2167
QY 2026 TCCACATTTAATGATGGATGGGCAATGCCAGTGCATGTGTAGGCAAGAAAGACCA 2085
Db 2168 TCCGCATTCAGATAAATGTTTTCGTACGCTAGTACCATGTTTCGTAAAAATAAGGCAGA 2227
QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGCCAGTAGAGCAAAACACCTGC 2145
Db 2228 CCAAGATAGTAACCTGATGAAGATTAAGAACTATGTAAGTAAGTGAGCCACTCACCC 2287
QY 2146 TGAGCCAGAAAG 2156
Db 2288 TGAATCTGATG 2298

RESULT 23
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
AC ABK15104;
XX
XX DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11-2.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
XX streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
XX Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
XX CDS 114..2630
XX /*tag= a
XX /product= "BVH-11-2"
XX /note= "The gene is flanked by sequences from the vector
XX SP64, no information on which is given in the
XX specification"
XX
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA000908.
XX
XX 20-JUN-2000; 2000US-0212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI: 2002-122272/16.
XX
XX P-PSDB; AAU75934.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
XX polypeptides, useful as vaccine components for treating or preventing
XX streptococcal infections such as otitis media, meningitis, and
XX bacteraemia.
XX
XX Example 3; Fig 5; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90% identity
XX to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
XX BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
```


CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae, or
 CC Streptococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumoniae protein BPH-11-2, used to create the antigenic
 CC peptides described in the method of the invention
 XX
 SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
 Query Match 41.1%; Score 980.8; DB 6; Length 2639;
 Best Local Similarity 67.8%; Pred. No. 2.6e-237;
 Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;
 QY 1 TTCTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
 DB 173 TTCTATGAACCTGGTGTGTCACCAAGCTGGTCAAGTTAAAGAGAGTCTAATCGAGTTTC 232
 QY 58 CTATATAGATGGAACACGACGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTAG 117
 DB 233 TTATATAGATGGTGTGTCACGAGTGGTCAAAAGGAGGAAATTTGACACCGAGTGAAGTCAG 292
 QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGT 177
 DB 293 TAAGAGAGAGGGATCAACGCGCAACAAATTTGATCAAGATTAACGATCAAGTTATGT 352
 QY 178 CACTTCAATGCGGACCACTATCATTAATTAATGATGTAAGGTTCTTATGAGCTATCAT 237
 DB 353 GACCTCTCATGAGACCAATTAATTAATTAATGATGTAAGGTTCTTATGATGCCATCAT 412
 QY 238 CAGTGNAGAAATTAATCAAGAGATCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
 DB 413 CAGTGNAGAAATTAATCAAGAGATCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
 QY 298 TGAGGTCAAGGGTGGGATATGTTATCAAGGTAGATGAAATTAATGTTTACCTTAAGGA 357
 DB 473 TGAATCAAGGGTGGGATATGTTATCAAGGTAGATGAAATTAATGTTTACCTTAAGGA 532
 QY 358 TCGTCCCAACGGGATAACGTCGTAACAAAGAGGAAATCAATTCGAACAAACAGAGCA 417
 DB 533 TCGGCGCCATGCGGACCAATTAATTCGACAAAGAGAGATTAACAGTCAAGAGCAGGAACA 592
 QY 418 TAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGTGTTGCCCTTGGCAGTTTC 477
 DB 593 CAGTCAATATCAT-----AACTCAAGAGCAGATAATGTTGTTGTCGACCCAGAGC 643
 QY 478 GCAAGGACGCTATCTACAGATGATGGTTATATCTTTAATGTTCTTGTATATCATAGAGGA 537
 DB 644 CCAAGGACGTTATACACCGATGATGGTATATCTTCAATGATCTGATATCATTTAGGA 703
 QY 538 TACTGTGATGCTTATATCGTTTCTCATGAGATCAATTAACATTAATTCCTTAAGAAATGA 597
 DB 704 CACGGTGTATGCTTATATCGTTTCTCATGAGGACCAATTAACATTAATTCCTTAAGAAATGA 763
 QY 598 GTTATCAGCTAGCGAGTGTGCTGTCAGAAAGCCTTCTTATCTGTTGTCGAGGAATCTGTC 657
 DB 764 GTTATCAGCTAGCGAGTGTGCTGTCAGAAAGCCTTCTTATCTGTTGTCGAGGAATCTGTC 798
 QY 658 AAATTCAGAAACCTATGCGCGACAAAATAGCGGATAACACTTCAAGAAACAAACTGGGTACC 717
 DB 799 -----ATTGAAATGGGAGCAGGGATCTCGTCTCTTTCAAGTTCTTAGTTATATGC 850
 QY 718 TTCTGTAAAGCAATCCAGGAATCAAAATATAACACAGCAACCAACAGCAACTAACAG 777
 DB 851 AAATCCAGTTCAACCAAGATTTGTCAGAGAAACCAATCTGACTGTCACTCCAACCTTATCA 910

QY 778 TCAAGCAAGTCAAGATATGACATTTGATAGTCTCTTTGAAAACAGCTCTTACAAAACGCTTT 837
 DB 911 TCA---AAATCAAGGGGAAAACATTTTCAAGCCCTTTTACGTGAATTTGATGCTAAACCCCT 967
 QY 838 GAGTCAACACATGTATGATCTGATGGCTTGTCTTTGATCCAGACAAATCAACAGTGC 897
 DB 968 ATCAGAACCCATGTATGATCTGATGGCTTTATTTTCGACCCAGCGCAATCAACAGTGC 1027
 QY 898 AACAGCTAGAGTGTGTCAGTCCACACGAGATCATTAACCACTTCCATCCCTTACTCTCA 957
 DB 1028 AACCGCAGAGGTAGTGTCCCTCATGTTAAACCACTTATCCACTTTTATCCCTTATGAACA 1087
 QY 958 AATGCTGAATTTGGAAGACGATCGCTCGTATTAATTCCTTCCCTTCTGTTATCGTTCAACCA 1017
 DB 1088 AATGCTGAATTTGGAAGAACGAATTCGTATTAATTCCTTCCCTTCTGTTATCGTTCAACCA 1147
 QY 1018 TTGGGTACCAGATTCAGGCCAGAACCAACAGTCCACAAACGACTCCGGAACCTAGTCC 1077
 DB 1148 TTGGGTACCAGATTCAGAACCCAGAACCAACAGTCCACAAATTCGAATTCGACTCCGGAACCTAGTCC 1207
 QY 1078 AGCCCGCAACCTGCACCAAAATCTTTAAATATAGACTCAAAATTTCTC-----TTT 1125
 DB 1208 AAGTCTGCAACCTGCACCAAAATCTTCAACAGCTCCCAAGCAATCCAATTTGATGAGAAAT 1267
 QY 1126 GCTTACTAGCTGTTACGAAAGTTGGGGAAGATATGTTTGAAGAAAGGGCATCTC 1185
 DB 1268 GGTCAAGAAAGTGTTCGAAAGATAGGCGATGGTTATGTTTGGAGGAAATGGAGTTTC 1327
 QY 1186 TCGTTATGTTTGGGAAAGATTTTAACTCTCAAACTGTTAAAAATCTTTGAAAGCAAGTT 1245
 DB 1328 TCGTTATATCCAGCCAGGATCTTTTACAGCAACACAGCAGGAGGATTTGATGACAACT 1387
 QY 1246 ATCAAAACAGAGAGTGTTCACACATTTTAACTGCTAAAAAGAAATTTGTTCTCTCG 1305
 DB 1388 GGCAAGCAGGAAAGTTTATCTCTAAGCTAGGAGCTAAGAAACCTGACCTCCCACTAG 1447
 QY 1306 TGACCAAGAAATTTATGATAAGCATATACTGTTTAACTGAGGCTCAATAAGCCCTGTT 1365
 DB 1448 TGATCGAGAAATTTTCAATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1507
 QY 1366 TGNAAATAGGGTCTGTAATTTCTGATTTTCAAGCCTTTAGACAAATTTATAGAACGCTTGA 1425
 DB 1508 TGATTAATAAGTTCGACAAAGTTGATTTTGAAGTTTGGATTAACCTGTTGGAACGACTCA 1567
 QY 1426 TGATGAATCGATTAATAAGAAAAATTTGATGATGATTTTGGCATTTCTAGCAACCAAT 1485
 DB 1568 GGAATGCTCAAGTGAATAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGCTCCGAT 1627
 QY 1486 TACCCATCCAGAGGACTTGGCAACCAAAATTTCTCAAAATTTGAGTATACCTGAGACGAA 1545
 DB 1628 TCGTATCCAGAACGTTTAGGAAACCAAAATTCGCAAAATTAACCTACACTGATGATGAT 1687
 QY 1546 TCGTATTTGCTCAATTAGCTGATAAGTATATAACAGTCAAGATGGTTTACATTTTGTGAAACA 1605
 DB 1688 TCAAGTAGCAAGTTTGGCAGGCAAGTACACACAGAGAGCGTTATATCTTTGATCCTCG 1747
 QY 1606 TGATATAATCAGTATGAAGAGATGATATGTAACGCTCATATGGGCGCATAGTCACTG 1665
 DB 1748 TGATATAACAGTATGAGGGGATGCTTATGTAATCTCACAATATGACCATAGCCACTG 1807
 QY 1666 GATTGGAAGAGATAGCTTCTGTAAGGAAAGTTGCGCTCAAGCTCAAGCTATACCTAAAGA 1725
 DB 1808 GATTAAAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCCAGGCTTATGCTAAAGA 1867
 QY 1726 AAAAGTATCTTACCTTCCATCTCCAGAGCAGATGTTAAAGCAAAATCCAATTTGGAGATAG 1785
 DB 1868 GAAAGTTTGACCCCTCTCTTCCAGAGACCAACAGGATTCAGGAAATCTGAGGCAAAAG 1927
 QY 1786 TGCAGCAGCTATTTTCAATTCGTTGTAAGGGGAAAAAGAAATTCACCTGTTCTGACTTCC 1845
 DB 1928 AGCAGAGCTATCTACACCGCGTGAAGAGCAGCTAAGAGGTGCCACTTGTATCGTATGCC 1987

1846 ATATATGTTGAGCATACAGTTCAGGTTAAACCGGTAATTTGATTCTCTATAGGA 1905
 1888 TTACAATCTCAATATACCTGTAGAGTCAAAACGGTAGTTTAAATCATACCTCATATGA 2047
 1906 TCATTACCAATAATTAATTTGCTTGGTTGATGATCACAATACAAAGCTCCAAATGG 1965
 2048 CCATTACCAATAACATCAATTTGAGTTGAGTTGAGGAGGCTTTATGAGGACCTAAGG 2107
 1966 CTATACCTTGAAGATTTGTTGGACGATTAAGTACTAGTAGACACCTGACGAACG 2025
 2108 GTATAGCTTGGAGATCTTTGGCGACTGTCAAGTACTATGTGCAACATCAACGAACG 2167
 2026 TCACATTTCAATGATCGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
 2168 TCCGCAATTCAGATAATGTTTGTGTAACGCTAGTGACCATGTTTCGTAAATAAGGCAGA 2227
 2086 CAGTGAAGATCCCAATTAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAACACCTGC 2145
 2228 CCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTGAAGTGAAGCACTCACCC 2287
 2146 TGAGCCAGAAAG 2156
 2288 TGAATCTGATG 2298

RESULT 24
 AAX25394
 ID AAX25394 standard; DNA; 2163 BP.

XX AC AAX25394;

XX DT 19-JUL-1999 (first entry)

XX DE Streptococcus pneumoniae complement C3-degrading protease DNA.

XX KM Human complement C3-degrading protease; vaccine; infection; meningitis;
 XX KW pneumonia; xerotransplantation; transplant rejection; inflammation; ds.

XX OS Streptococcus pneumoniae.

XX PN W0915675-A1.

XX PD 01-APR-1999.

XX PF 24-SEP-1998; 98WO-US020186.

XX PR 24-SEP-1997; 97US-0059907P.

XX PA (MINU) UNIV MINNESOTA.

XX PA (AMCY) AMERICAN CYANAMID CO.

XX PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;

XX DR WPI; 1999-254719/21.

XX DR P-PSDB; AAY05753.

XX PT New isolated human complement C3-degrading proteinase.

XX PS Claim 54; Page 52-54; 66pp; English.

XX CC This DNA sequence encodes a 79 kDa protein (see AAY05753) of
 CC Streptococcus pneumoniae serotype 4 that is capable of degrading human
 CC complement protein C3 (HCP3). It was identified in the S. pneumoniae
 CC serotype 4 genome by alignment to another novel open reading frame (see
 CC AAX25393) that codes for a 20 kDa HCP3 protease (AAY05752) of S.
 CC pneumoniae serotype 3. This suggested the open reading frame that codes
 CC for the 20 kDa protein may be part of a larger open reading frame. Amino
 CC acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence
 CC identity with amino acids 170-227 and 258-300 of the 79 kDa protein.
 CC proteins and peptides or polypeptides containing these regions, and DNA
 CC sequences encoding them (nucleotides 507-681 and 827-999 of the present
 CC sequence) are claimed. HCP3 proteases and polypeptides can be used as
 CC immune system stimulating compositions (claimed). They can produce an

CC immune response against S. pneumoniae to immunize or treat a mammalian
 CC subject against infection or colonization (claimed). They can produce a B
 CC cell response, a T cell response, an epithelial cell response, or an
 CC endothelial cell response (claimed). The expression of the proteins on
 CC the surface of an organ of an animal used in xenotransplantation can be
 CC used to inhibit C3-mediated inflammation and rejection
 XX SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 U; 0 Other;

Query Match 36.4%; Score 869.4; DB 2; Length 2163;
 Best Local Similarity 65.2%; Pred. No. 3.5e-209;
 Matches 1436; Conservative 0; Mismatches 627; Indels 141; Gaps 5;
 QY 254 ATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTTATGAGGTCAAGGGTGA 313
 DB 1 ATGAAGATCCGAATATATCATGTTCAAGGATTCAGACATGTTCAATGAATCAAGGGTGT 60
 QY 314 TATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGTCGCCAGCGGAT 373
 DB 61 TATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGTCAGCTCATGCCGAT 120
 QY 374 AACGTCGTCACAAAGAGGAAATCAATCGACAAACAAAGAGCATAGTCAACATCGTGAA 433
 DB 121 AATATTCGACAAAGAGAGATTAAACGTCAGAGCGGAAACACAGTCTATATACGGG 180
 QY 434 GGTGGAATCTCCAAGAAACGATGGTGTGTTGCTTGGCAGCTTGGCAAGGACGCTATACT 493
 DB 181 GGTGTTCT-----AACGATCAAGCAGTAGTTGCGCCAGAGGCCAAGCAGCTATACA 234
 QY 494 ACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATCTGGTGATGCTTAT 553
 DB 235 ACGGATGATGGTTATATCTTTCAATGCTATCTGATATCATTTGAGGACACGGGTGATGCTTAT 294
 QY 554 ATGTTCTCTCATGAGATCATACCATTCATCTTAAGATGATGTTATCAGCTAGCGAG 613
 DB 295 ATGTTCTCTCAGCGGACCATTCACATTCATCTTAAGATGATGTTATCAGCTAGCGAG 354
 QY 614 TTGGCTCTCAGAAAGCCTTCTCTATCTGTCGAGGAAATCTGTCAAATTCAAAGAACCTAT 673
 DB 355 TTAGTCTCTCAGAAAGCCTTAT-----GGAAAT 381
 QY 674 GCGGCAAAATAGCGATACACATTCAGAAACAAACTGGGTACCTTCTGTAAAGCAATCCA 733
 DB 382 GGGAAAGCAGGGATCTCGTCTCTTCAAGTTCTAGTTTATATGCAAAATCCAGCTCAACCA 441
 QY 734 GGAATACAAATCTAACACAAAGCAACACAGCAACTAACAGTCAAGCAAGTCAAAGT 793
 DB 442 AGATTGTCAGAGAACCAATCTGACTGTCTACTCAACTTA---TCATCAAAATCAAGGG 498
 QY 794 AATGACATTTGATAGTCTCTTGAACACAGCTCTACAAACTGCTTTGAGTCAACGACATGTA 853
 DB 499 GAAACACATTTCAAGCCTTTTACGTTGAATTTGATGCTAAACCCCTTATCAGAACGCGCATGTG 558
 QY 854 GAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAGAGTCCGACAGCTAGAGGTGTT 913
 DB 559 GAATCTGATGGCTTATTTTTCGACCCGCAATCAAGTCCGACAGGAGGTGTA 618
 QY 914 GCAGTGCACACGGAGATCATTAACCACTTCTCCCTTACTCTCAAATGTCTGAATTTGAA 973
 DB 619 GCTGTCCCTCATGTAACCATTAACCATTTATCCCTTATGAAACAAATGTCGAATTTGAA 678
 QY 974 GAACGAATTCGTCGTATTTATTCCTTCTGTTATGTTTCAACCAATGGGTACCAAGATTC 1033
 DB 679 AAACGAATTCGTCGTATTTATTCCTTCTGTTTCAAAACCATTTGGGTACCAAGATTC 738
 QY 1034 AGGCAGAAACCAAGTCCACACCCAGTCCGGAACCTAGTCCAGCGCCGCAACCTGCA 1093
 DB 739 AGACAGAAACCAAGTCCACATTCGACTCCGGAACCTAGTCCAGTCCGCAACCTGCA 798
 QY 1094 CCAATCTTTAAATAGACTCAA-----ATTCTTTCTTTGGTTAGTCACTGGTA 1141
 DB 799 CCAATCTCAACCCAGCTCCAAGCAATCCAATTCATGAGAAATTTGGTCAAAAGAGCTGTT 858

1142 CGAAAAAGTTCGGGAAGGATATGTTATCGAAGAAAAGGCGATCTCTCGTTATGCTTTGGC 1201
 859 CGAAGAGTAGGCGATGGTTATGCTTTGAGAGAGATGGAGTTCTCGTTATATCCAGCC 918
 1202 AAGAATTACATCTGAACTGTAAATCTTTGAAAGCAAGTTATCAAAACAGAGAGT 1261
 919 AAGGATCTTTGAGT 978
 1262 GTTTCACACACTTTAACTGCTAAAGAAAAGAAATGTTGCTCTCGTACCAAGAAATTTAT 1321
 979 TTAATCTCAAGCTAGAGAGCTAGAGAACTGACCTCCCATCTAGTGCATCGAGAAATTTAC 1038
 1322 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGAGCTTTGTTGNAATAAGGTCGT 1381
 1039 AATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGATATAAGGTGGA 1098
 1382 AATCTGATTTCAAGCCTTTAGACAAATTTATGAAAGCTTTGAATGATGAATCGACTAAT 1441
 1099 CAAGTTGATTTGAGGCTTTGGATAACCTGTTGGAAGCACTCAAGGATGTCCCAAGTAT 1158
 1442 AAGAAGAAATTTGATAGATGTTTATTTGGCATTTCTAGCACCATTACCCATCCAGAGCGA 1501
 1159 AAGTCAAGTTAGTGGATGATATCTTTGGCTTTCTTAGCTCCGATTCGTATCCAGAACGT 1218
 1502 CTTGGCAAAACCAAAATCTCAAAATGAGTATGATGAAAGCAAGTTCGTATTGCTCAATTA 1561
 1219 TTAGGAAACCAAAATCGCAAAATTTACTACATGATGAGATTCAGTAGCCAAGTTG 1278
 1562 GCTGATAAGTATACAAAGCTGAGATGTTTATTTGATGAAATCATGATTAATCAAGTAT 1621
 1279 GCAGGCAAGTACACAAAGAGAGAGGTTATCTTTGATCTCGTATTAACCAAGTAT 1338
 1622 GAAGAGATGATATGATGAGGCTCATATGGGCACTAGTCACTGGATGGAAAGATAGC 1681
 1339 GAGGGGATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 1682 CTTTCTGATAAGGAAAAGTTCAGCTCAAGCTATGATGATGATGATGATGATGATGATGAT 1741
 1399 TTGCTGAGCTGAGT 1458
 1742 CATCTCCAGAGAGATGTTAAAGCAATTCCTGAGATGATGATGATGATGATGATGATGAT 1801
 1459 CTTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1518
 1802 AATCGTGTGAT 1861
 1519 AAGCGGTGAAGAT 1578
 1862 ACAGTTGAGGTTAAAAACGGTAAATTTGATTTCTCTCAATGATGATGATGATGATGATGAT 1921
 1579 ACTGTAGAGTCAAAACGGTAGTTTAATCATACCTCATTTATGACCATTTACCATACATC 1638
 1922 AATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981
 1639 AATTTGAGTGTGTTGAT 1698
 1982 TTGTTTGGCAAGATTAAGTACTAGTACGATGATGATGATGATGATGATGATGATGATGAT 2041
 1699 CTTTGGGCACTGTCAAGTACTATGTCGAACTATGTCGAACTATGTCGAACTATGTCGAA 1758
 2042 GGATGGGCAATGCCAGTGAATGTTGTTAGGCAAGAAAGACCAAGTGAAGATCAAAAT 2101
 1759 GGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1818
 2102 AAGAATTTCAAGCGGATGA-----AGAGCCAGTAGAG 2134
 1819 CAAACGGAAGAAACCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
 2122 -----AGAGCCAGTAGAG 2134
 1879 CGAAGAGAGAAACCGCAAG 1938
 2135 GAAACACCTGCTGAGCCAGAGAGTCCCTCAAGTAGAGAGAGAGAGAGAGAGAGAGAGAG 2194

1939 GAATCACCAGAGAGATCAGAAAGAACTCAGGTCGAGACTGAAAGGTTGAAGAAAACCTG 1998
 2195 AAAGAAGCAGAGAGATTTTGTCTTGGGAAAGTAAACGGATTTCTAGTCTGAAAGGCAATGCAACA 2254
 1999 AGAGAGGCTGAAGATTTTACTTGGGAAATCCAGGATCCAAATTTCAAGTCCATGCCAAA 2058
 2255 GAAACTCTAGCTGGTTTACGAAATTAATTTGACTCTTCAAAATTTGATGATGATGATGATGAT 2314
 2059 GAGACTCTCAGAGATTTAAATAATTTACTATTTGGCACCAGGACCAACATACTATT 2118
 2315 ATGGCAG 2358
 2119 ATGGCAG 2162

RESULT 25
 AAV52376
 ID AAV52376 standard; DNA; 2359 BP.
 XX AAV52376;
 AC AAV52376;
 XX 23-OCT-1998 (first entry)
 DT 23-OCT-1998 (first entry)
 XX Streptococcus pneumoniae genome fragment SEQ ID NO:243.
 DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 XX computer readable medium; vaccine; pharmaceutical composition; ds.
 KW Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX WO9818931-A2.
 PN 07-MAY-1998.
 XX 30-OCT-1997; 97WO-US019588.
 XX 31-OCT-1996; 96US-0029960P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
 PI Dougherty BA;
 XX WPI; 1998-272225/24.
 XX Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
 XX Claim 1; Page 1265-1266; 1409pp; English.
 PS The present invention describes a computer readable medium which has the
 CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
 CC 391, identifying members of the library which contain sequences that
 CC hybridise to the target sequence and isolating the nucleic acid molecules
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
 CC homologous to amplification primers derived from the fragment of the S.
 CC pneumoniae genome to prime the amplification and isolating the amplified
 CC sequences. The computer readable medium can be used in a computer-based
 CC system for identifying fragments of the S. pneumoniae genome of
 CC commercial importance, or expression modulating fragments of the S.
 CC pneumoniae genome. Products from the present invention can be used in
 CC diagnosis kits and assays, and pharmaceutical compositions and vaccines

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 11:19:10 ; Search time 6082 Seconds
(without alignments)
11729.817 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	3.0	906	28	BH153606
2	69.8	2.9	890	28	BH146886
3	67.8	2.8	180	28	AF325819
4	67.8	2.8	908	28	AZ548467

					BH153606
					BH146886
					AF325819
					AZ548467

C	5	66.4	2.8	829	29	BX173672
C	6	64.8	2.7	843	28	AZ551618
C	7	63.8	2.7	877	28	AZ531291
C	8	62.2	2.6	912	28	AZ551092
C	9	60.8	2.5	849	28	AZ546009
C	10	59.2	2.5	931	28	BH160272
C	11	58.6	2.5	1101	29	CNS0039G
C	12	57.2	2.4	900	28	AZ549980
C	13	56.8	2.4	905	28	AZ550256
C	14	56.6	2.4	645	12	BM165350
C	15	56.6	2.4	747	12	BM162732
C	16	55.2	2.3	816	28	AZ535744
C	17	55	2.3	671	12	BM163120
C	18	55	2.3	647	12	BM160252
C	19	55	2.3	694	12	BM168242
C	20	55	2.3	717	12	BM160500
C	21	55	2.3	756	12	BM162492
C	22	55	2.3	794	12	BM159906
C	23	55	2.3	795	28	AZ528485
C	24	55	2.3	823	28	AZ676218
C	25	55	2.3	843	28	BH139532
C	26	53.8	2.2	721	13	BM141179
C	27	53.6	2.2	656	13	BM144040
C	28	53.6	2.2	844	29	BX139987
C	29	53.2	2.2	752	14	CD099026
C	30	52.8	2.2	671	13	BW273704
C	31	52.8	2.2	890	28	AZ530768
C	32	52.4	2.2	677	12	BM161314
C	33	52.4	2.2	725	12	BM169486
C	34	52.4	2.2	866	28	AZ527885
C	35	52.4	2.2	976	28	BH149983
C	36	52.2	2.2	945	29	CNS008CA
C	37	51.8	2.2	435	12	BM169710
C	38	51.8	2.2	880	28	AZ529191
C	39	51.6	2.2	853	29	CNS023KH
C	40	51	2.1	500	9	AU088119
C	41	50.8	2.1	726	28	AZ196050
C	42	50.2	2.1	1007	29	CNS06X9S
C	43	50	2.1	445	12	BM167469
C	44	50	2.1	555	13	BU496963
C	45	49.8	2.1	707	12	BM164483
C	46	49.6	2.1	598	12	BM170666
C	47	49.6	2.1	605	12	BM163520
C	48	49.4	2.1	629	12	BW073162
C	49	49.4	2.1	687	12	BM169665
C	50	49	2.1	1044	13	BX415231
C	51	48.8	2.0	1101	29	CNS0022U
C	52	48.6	2.0	605	12	BM162501
C	53	48.6	2.0	861	13	BW266138
C	54	48.4	2.0	546	13	BU497202
C	55	48.4	2.0	891	28	AZ683582
C	56	48.4	2.0	960	28	CC143364
C	57	48.2	2.0	468	12	BG553005
C	58	48.2	2.0	537	13	BU498263
C	59	48.2	2.0	554	12	BM275335
C	60	48.2	2.0	648	10	BE349803
C	61	48.2	2.0	661	9	AU060703
C	62	48.2	2.0	679	12	BJ440521
C	63	48.2	2.0	756	12	BM165937
C	64	48.2	2.0	782	13	BM117562
C	65	48	2.0	548	12	BM159818
C	66	48	2.0	1101	29	CNS00LT2
C	67	47.8	2.0	713	13	BU412428
C	68	47.8	2.0	718	13	BU257897
C	69	47.8	2.0	773	13	BU249618
C	70	47.8	2.0	792	13	BU114945
C	71	47.8	2.0	799	13	BU332339
C	72	47.8	2.0	834	12	BU090507
C	73	47.8	2.0	884	13	BU137322
C	74	47.6	2.0	880	28	AZ669474
C	75	47.4	2.0	541	29	CNS04KUS
C	76	47.4	2.0	731	13	BM115605
C	77	47.2	2.0	311	14	CA699402

BX173672 Danio rer
AZ551618 ENTDS4TR
AZ531291 ENTBO34TR
AZ551092 ENTBJ22TF
AZ546009 ENTFW53TF
BH160272 ENTQV49TR
AL063921 Drosophila
AZ550256 ENTEV58TR
BM165350 EST567873
BM162732 EST565255
AZ535744 ENTQ25TR
BM163120 EST565643
BM160252 EST562775
BM168242 EST570765
BM160500 EST563023
BM162492 EST565015
BM159906 EST562429
AZ528485 ENTCM64TF
AZ676218 ENTNG88TF
BH139532 ENTNG88TF
BM141179 BM141179
BM144040 BM144040
BX139987 Danio rer
CD099026 AGENCOURT
BW273704 BW273704
AZ530768 ENTBH54TF
BM161314 EST563837
BM169486 EST572009
AZ527885 ENTCA79TR
BH149983 ENTQ93TF
AL051492 Drosophila
BM169710 EST572233
AZ529191 ENTBV68TR
AL179594 Tetraodon
AU088119 AU088119
AZ196050 SP_1031_A
AL19462 T3 end of
BM167469 EST56992
BU496963 PESTOab6
BM164483 EST567006
BM170666 EST573189
BM163520 EST566043
BM169665 EST572188
BJ073162 BJ073162
BX415231 BX415231
AL097152 Drosophila
BM162501 EST565024
BW266138 BW266138
BU497202 PESTOab5
AZ683582 ENTKK47TR
CC143364 ND.L.32A13
BG553005 dab82e09
BU498263 PESTOab9
BM275335 PESTOab6
BE349803 hq43c11.x
AU060703 AU060703
BJ440521 BJ440521
BM165937 EST568460
BM117562 BM117562
BM159818 EST562341
AL078714 Drosophila
BU412428 603154380
BU257897 603745789
BU249618 603778405
BU114945 603131025
BU332339 603500448
BU090507 BJ090507
BU137322 603122583
AZ669474 ENTYH88TR
AL296309 Tetraodon
BM115605 BM115605
CA699402 wlk8.pk00

DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence s
High quality sequence s

FEATURES

source

```

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica"
/notes="vector: pHOS1; Site 1: Bst
Institute for Genomic Research (TIGR)
Genomic DNA isolated from broth
culture of HMI:IMSS. The library
using a method described by Clark,
C.G. and Diamond, L.S. (1993) EMB
method for isolate identification
77:450). The DNA was mechanically
tight size distribution (~2 kb).
The library construction is described
H.O. and Venter, J.C. (Making sm
whole genome shotgun sequencing
Sequencing: A Practical Approach,
Baxell, Oxford University Press,

```

ORIGIN

Query Match 2.9%; Score 69.8; DB 28; Length 890;
Best Local Similarity 43.6%; Pred. NO. 1e-05;
Matches 311; Conservative 0; Mismatches 402; Indels 0

1497	QY	AGCGACTTGGCAACCAAATCTCMAATTGAGTATACTGAAGACGAAGTTGCTGATTGCTC	1556
15	Db	AGAGATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAGAAGATGATGAAG	74
1557	QY	AATTAGCTGATAAGTATACAACGTTCAGATGGTTACATTTTTTGATGAACATGATATATCA	1616
75	Db	ATCACGATGATGATGAAGATGCAGATGATCATGNAGNAGATGATGCAGATGAAGAAG	134
1617	QY	GTGATGAAGAGATGCATATGTAAAGCCCTCATATGGGCCCATAGTCACTGGAATGGAAAAAG	1676
135	Db	ACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAG	194
1677	QY	ATAGCCCTTTCTGATTAAGGAAAAGTTCCAGCTCAAGCCTATCTAAGAGAAAAGGTTATCC	1736
195	Db	ACGATGAAGAAGACGATGAAGAAGACGATGATGAAGAAGACGATGAATTTGAATTAG	254
1737	QY	TACCTCCATCTCAGACGCAGATGTTAAAGCAAAATCCAATGGAGATAGTGCAGCAGCTA	1796
255	Db	AAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAG	314
1797	QY	TTTTACAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCCACTTCCATATATGTTG	1856
315	Db	AAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAG	374
1857	QY	AGCATACAGTTGAGTTTAAAAACGGTAAATTTGATTTATCTCTAAGAGATCATCAATA	1916
375	Db	AAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAG	434
1917	QY	ATATTAAATTTGCTTGGTTTGATGATCACACATACAAAAGCTCCAAATGGCTATACCTTGG	1976
435	Db	AAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAG	494
1977	QY	AAGATTTGTTGGACGATTAAGTACTACTGTAGAACACCCCTGACGAACGTCCACATTTCTA	2036
495	Db	AAGATGAAGATGATGAAGATGATGATGAAGAAGACGATGAAGAAGATGATGAAGAAGACG	554
2037	QY	ATGATGATGGGCAATGCCAGTGAGCATGTGTTAGCAGAAGAACCCACAGTGAAGATC	2096
555	Db	ACGATGAAGACGATGAAGACGATGAAGACGATGAAGACGATGAAGACGATGAAGAAG	614
2097	QY	CAAAATGAACATTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCCTGCTGAGCCAGAAG	2156

Db 615 ACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAG 674

Qy 2157 TCCTCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAAGCAGAACTT 2209

Db 675 ACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGAT 727

RESULTS

AF325819	180 bp	DNA	linear	GSS 08-MAR-2002
AF325819	Leptospira interrogans serovar lai	DNA	Leptospira	
interrogans serovar lai	genomic clone 18,		genomic survey	sequence.
AF325819				
AF325819.1	GI:19263459			
GSS.				
Leptospira interrogans serovar lai				
Leptospira interrogans serovar lai				
Bacteria: Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.				
1 (bases 1 to 180)				
Ku, C. and Bao, H.				
Leptospira interrogans serovar lai	DNA			
Unpublished (2002)				
Contact: Hu C				
Leptospirosis Research Unit				
West China University of Medical Sciences				
17# Renming Nan Road, Chengdu, Sichuan 610041, P.R. China				
Email: huchanghua@263.net				
Class: unknown.				

FEATURES

source

```
1.180
/organism="Leptospira interrogans serovar lai"
/mol_type="genomic DNA"
/db_xref="taxon:57678"
/clone="18"
/clone_lib="Leptospira interrogans serovar lai
/note="obtained through DNA subtraction using
nonpathogenic strain Leptospira biflexa serovar
strain Patoc I"
```

ORIGIN

	Query Match	2.8%	Score 67.8	DB 28	Length 180
	Best Local Similarity	62.9%	Pred. No. 1.1e-05		
	Matches 105	Conservative 0	Mismatches 62	Indels 0	Gaps 0
QY	249	TACTCATGGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTATGAGGTCAAGG	308		
Db	1	TATTAATGACGGACCCCAATTATCAGTTTAAACAGTCTGATGTTGTTAAACGAGATTCTAG	60		
QY	309	GTGGGATATGTTTCAAGGTAGATGGAAAACTACTATGTTTACCTTAAGGATGTCGCCACG	368		
Db	61	ATGGCTATATTATCAGGTGTTGATGGCAAGTATTATGTTTATCTTAAGCCAGGAGCAAGC	120		
QY	369	CGGATAACTCCGTCACAAAGAGGAAATCAATCGACAAAAACAAGAG	415		
Db	121	GTAAAAAATTCGTCACCAAGCAGCAATTCGTGAACAGGTTTGAAG	167		

RESULT 4

Accession	AZ548467	908 bp	DNA	linear	GSS 14-NOV-2000
Definition	ENT3K30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.				
Version	AZ548467.1	GI:11172102			
Source	GSS.				
Keywords	Entamoeba histolytica				
Organism	Entamoeba histolytica				
Reference	Eukaryota; Entamoebidae; Entamoeba.				
Authors	1 (bases 1 to 908)				
Title	Lofthus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica				
Journal	HM1:IMSS sheared DNA library Unpublished (2000)				

COMMENT

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 828.

FEATURES

source

1..908
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="vector: pHOS1; Site 1; Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

ORIGIN

Query Match 2.8%; Score 67.8; DB 28; Length 908;
 Best Local Similarity 44.0%; Pred. No. 3.2e-05;
 Matches 334; Conservative 0; Mismatches 422; Indels 3; Gaps 1;
 QY 1525 TGAGTATCTGAAGACGAGTGGTATTGGCTCAATTAGCTGATAGTATACAACTGACGA 1584
 DB 33 TGAAGATGATGATGATGAAGAAGATGATGATGATGAAGAAGATGATGATGAAGAAGATGATGA 92
 QY 1585 TGGTTACATTTTTCATGAACATGATATATCAATCAGTGAAGAGATGATGATGAAGCC 1644
 DB 93 TGATGAAGACATGATGAAGAAGATGATGAAGAAGACAAAGATGATGAAGATGAAGATGA 152
 QY 1645 TCATATGGCCATGATCTACTGGATTG---GAAAGATAGCCCTTCTGTGATAAGGAAAAGT 1701
 DB 153 TGAAGACGATGAAGACGACGATATGAATTAGACATGATGATGATGAAGAAGACGAAGA 212
 QY 1702 TGCAGCTCAAGCCTTACTTAAGAAAAGGTATCTTACTCCATCTCCAGACGACGATGT 1761
 DB 213 TGATGATGAAGATGATGAAGACGATGAAGACGACGATATGAATTAGAAAGATGATGATGA 272
 QY 1762 TAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTCAATCGTGTGAAGGGGAAA 1821
 DB 273 TGATGAAGACAGATGATGATGAAGAAGACAGATGATGATGAAGATGATGATGAAGA 332
 QY 1822 ACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCATCAGTTGAGGTTTAAACGG 1881
 DB 333 AGAAGATGAAGAAGATGATGAAGAAGATGATGAAGATGATGATGAAGAAGACGATGA 392
 QY 1882 TAAATTTGATTTCCTCATAGGATCAATACCATATATTAATTTGCTTGGTTTGATGA 1941
 DB 393 TGATGAAGATGATGAAGACGATGAAGACGACGATATGAATTAGAAAGATGATGAAGA 452
 QY 1942 TCACACATACAAAGCTCCAAATGGCTATACCTTGGAAAGATTTGTTTGGCAGCATTAAGTA 2001
 DB 453 ACAAGAAGATGATGATGAAGAAGACGATGATGATGAAGATGATGATGAAGAAGACGA 512
 QY 2002 CTACCTAGAACACCTTGACGAACGCTCCACATTTCTAATGATGATGGGCAATGCCAGTGA 2061

DB 513 TGAAGATGATGATGAAGAAGACGATGAAGATGATGATGAAGAAGACGATGATGATGAAGA 572
 QY 2062 GCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGA 2121
 DB 573 AGAAGATGAAGAAGACGATGATGAAGAAGATGATGAAGAAGACGATGATGATGAAGA 632
 QY 2122 AGAGCCAGTAGAGAAACACCTCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAAGT 2181
 DB 633 AGACGACGAATATGAATTAGAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGA 692
 QY 2182 AGAAGCCCAACTCAAGAAGCAGAGATTTTCTTGGCAAGATTAACGATTTCTAGTCTGAA 2241
 DB 693 AGATGATGAAGAAGACGATGATGATGATGATGATGAAGAAGATGATGATGAAGAATGA 752
 QY 2242 AGCCATGCAACACAGAACTCTAGCTGGTTTACGAAATAA 2280
 DB 753 TGAAGAAGATGAAGATGATGAAGAAGATGATGAAGAAGATGA 791

RESULT 5

BX173672/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..829

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="DKEY-150M6"

/tissue_type="Testis"

/note="vector pIndigoBAC-536"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;


```
High quality sequence start: 22
High quality sequence stop: 829.
Location/Qualifiers
1. .877
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="vector: pHS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
2.7%; Score 63.8; DB 28; Length 877;
Query Match 2.6%; Score 62.2; DB 28; Length 912;
Best Local Similarity 43.4%; Pred. No. 0.00029;
Matches 293; Conservative 0; Mismatches 382; Indels 0; Gaps 0;
QY 1535 GAAGCAGAGTTCGATTGCTCAATTAGCTGATAGTATACACGTCAGATGGTTACATT 1594
DB 785 GAAGAAGATGATGATGATGATGAAGAAGAGATGATGAAGATGATGATGAAGAA 726
QY 1595 TTTGATGAACATGATATATCACTGATGAAGAGATGATGAACGCTCATATGGGC 1654
DB 725 GATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAGAAGATGATGAAGAA 666
QY 1655 CATAGTCACTGGAATTGGAAGAGATAGCCTTTCTGATAGGAAAAAGTTGACGCTCAAGCC 1714
DB 665 GATGATGATGAAGAAGAGATGATGAAGATGAAGATGATGAAGAGATGAAGAGAA 606
QY 1715 TATCTAAGAAAAAGGATGATCCTACCTCCATCTCCAGCGCAGATGTTAAAGCAATCCA 1774
DB 605 TATGAATTAGAAGATGATGATGATGAAGAAGAGATGATGATGAAGATGATGAAGAC 546
QY 1775 ACTGAGATGTCGAGCGCTATTACATCTGTGAAGGGGAAAAAGTTGACGCTCAAGCTC 1834
DB 545 GATGAAGACGACGATATGAATTAGAAGATGAAGATGATGAAGAAGAGATGATGAT 486
QY 1835 GTTCGACTTCCTCATATATGTTGAGCATACAGTTGAGGTTAAAAAGCGGTAATTTGATT 1894
DB 485 GAAGAAGAGAGATGATGATGAAGATGATGATGAAGAAGAGATGAAGAAGATGATGAA 426
QY 1895 CCTATAAGGATCAATACCATATATTAATTTGCTTGGTTGATGATCAGATACATACAAA 1954
DB 425 GAAGATGATGAAGATGATGATGAAGAAGAGATGATGATGAAGATGATGAAGACCAT 366
QY 1955 GCTCAATGGTATACCTTGGAAGATTTGTTGACGATTAAGTACTACTAGTAAACAC 2014
DB 365 GAAGACGACGATATGAATTAGAAGATGATGATGAAGAAGAGAGATGATGAAGAA 306
QY 2015 CCTGACGACGCTCCACATTTCTTAATGATGGATGGGCAATGCCAGTGAGCATGTGTTAGGC 2074
DB 305 GAAGAAGATGATGATGAAGATGATGATGAAGAAGAGATGAAGATGATGAAGAAGAA 246
QY 2075 AAGAAGACCAAGTGAAGATCCAAATAGAAGCTTCAAGCGGATGAAGGCCAGTAGAG 2134
DB 245 GATGAAGATGATGATGAAGAAGAGATGATGATGAAGAAGAGAGATGAAGAAGAGAT 186
QY 2135 GAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTC 2194
DB 185 GATGAAGAGATGATGAAGAAGAGAGATGATGATGAAGAGAGATGAAGATGAATGAAGAA 126
QY 2195 AAAGAAGCAGAAAGTT 2209
```

```
Db 125 GATGATGAAGAGAT 111
| | | | | | | | | |
RESULT 8
AZ551092 912 bp DNA linear GSS 14-NOV-2000
LOCUS ENTJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.
ACCESSION AZ551092
VERSION AZ551092.1 GI:11176393
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 912)
Loftus,B., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library
Unpublished (2000)
JOURNAL
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@loftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 861.
FEATURES
Location/Qualifiers
1..912
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="vector: pHS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
ORIGIN
Query Match 2.6%; Score 62.2; DB 28; Length 912;
Best Local Similarity 43.3%; Pred. No. 0.00071;
Matches 292; Conservative 0; Mismatches 383; Indels 0; Gaps 0;
QY 1535 GAAGCAGAGTTCGATTGCTCAATTAGCTGATAGTATACACGTCAGATGGTTACATT 1594
DB 213 GAAGAAGATGATGATGATGATGAAGAAGAGATGATGAAGATGATGAAGATGATGAAGAA 272
QY 1595 TTTGATGAACATGATATATCACTGATGAAGAGATGATGAATGTATGTAAACCCCTCATATGGGC 1654
DB 273 GATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAGAAGATGATGAAGAA 332
QY 1655 CATAGTCACTGGATTGGAAGAGATAGCCCTTCTGATAGGAAAAAGTTGACGCTCAAGCC 1714
DB 333 GATGATGAAGAAGAGAGATGATGAAGATGAAGATGATGAAGATGAAGACGACGAA 392
QY 1715 TATCTAAGAAAAAGGATGATCCTACCTCCATCTCCAGACGCGAGATGTTAAAGCAATCCA 1774
```

Db 393 TATGAATTAGAAGATGATGATGATGAAGAAGAAGATGATGATGAAGATGATGAAGAC 452
 QY 1775 ACTGGAGATAGTCACGAGCTATTTCATCTGTTGAAGGGGAAAAACGAATTCACCTC 1834
 Db 453 GATGAAGACGACCAATATGAATTTAGAAGATGATGATGATGAAGAAGAAGATGATGAT 512
 QY 1835 GTTCGACTTCCATATATGTTTGAGCATACAGTTGAGGTTTAAACCGTAAATTTGATT 1894
 Db 513 GAAGAAGAAGAGATGATGATGAAGATGATGATGAAGAAGAAGATGATGAAGAAGATGA 572
 QY 1895 CTTCTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1954
 Db 573 GAAGATGATGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGA 632
 QY 1955 GCTCCAAATGGCTATACCTTGGAGATTTCTTGGACGATTAAGTACTACGTAGAAC 2014
 Db 633 GAAGACGACGAATGAAATTTAGAAGATGATGATGAAGAAGAAGATGATGATGAAGA 692
 QY 2015 CTTGACGAACGCTCCACATTTCTAATGATGGATGGGCAATGCCAGTCAGCATGT 2074
 Db 693 GAAGAAGATGATGATGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAA 752
 QY 2075 AGAAGAACCACAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2134
 Db 753 GATGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGAAGAAGA 812
 QY 2135 GAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGMAAAGTAGAAGCCCACT 2194
 Db 813 GATGAAGAAGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGAAGAAG 872
 QY 2195 AAAGAAGCAGAAGTT 2209
 Db 873 GATGATGAAGAAGAT 887

RESULT 9
 A2546009/c
 LOCUS ENTFW53TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2546009
 VERSION A2546009.1 GI:11167130
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 849)
 Loftus,B., Van Aken,S. and Fraser,C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)
 JOURNAL Contact: Brendan J Loftus
 COMMENT Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Claes: shotgun
 High quality sequence start: 26
 High quality sequence stop: 796.
 FEATURES
 source Location/Qualifiers
 1..849
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHO51; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Query Match 2.5%; Score 60.8; DB 28; Length 849;
 Best Local Similarity 43.4%; Pred. No. 0.0015;
 Matches 275; Conservative 0; Mismatches 358; Indels 0; Gaps 0;
 QY 1190 TATGCTTTTGGCAAGATTTACCATCTGAAACTGTTTAAAAATCTTGAAGCAAGTTATCA 1249
 Db 742 TTGACTTAGTGAATAATGTAGAAATGGAACATGATACCGTTAGAAGACATGAATTA 683
 QY 1250 AAACAAGAGAGTGTTCACACACTTTTAAGTGTCTAAAAAGAAAAATGCTCTCCTCGTAC 1309
 Db 682 GAAGAAGATGATGATGAATTTGAAATTTAGAAGACGAAGATGATGATGATGATGAA 623
 QY 1310 CAAGAAATTTATGATAAGCATATAATCTGTTAACTGAGGCTCAAAAGCCTTTGTTGNA 1369
 Db 622 GACGAAGAAGATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
 QY 1370 AATAAGGGTGGTAAATTTCTGATTTCCAGGCTTTAGACAAATTTATAGAACGCTTGAATGAT 1429
 Db 562 GAATTTGAATTTAGAAGATGAAGAAGATGAAGATGATGATGATGATGATGATGATGATGAT 503
 QY 1430 GAATCGACTTAATAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCATTACC 1489
 Db 502 GAAGATGATGATGAAGATGAAGAAGATGATGATGATGATGAAGAAGACGAAGAAGATGAT 443
 QY 1490 CATTCAAGAGCAGCTTGGCAACCAATTTCTAAATTTAGTATGATGATGATGATGATGATGAT 1549
 Db 442 GATGATGATTTGAAATTTAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGATGATGAT 383
 QY 1550 ATTGCTCAATTAGCTGATAAGTATACACGTCAGATGTTTACATTTTGTGATCAACATGAT 1609
 Db 382 GATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGAT 323
 QY 1610 ATAATCAGTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
 Db 322 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 263
 QY 1670 GGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729
 Db 262 GAAGATGAAGAAGATGAAGAAGATGATGATGAAGAAGATGAAGAAGATGAAGAAGATGATGAT 203
 QY 1730 GGTATCTTACCTCCATCTCCAGACGATGTTAAAGCAAAATCCAACTGGAGATGATGCA 1789
 Db 202 GAATTAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGAA 143
 QY 1790 GCAGCTATTTCAATCGTGTGAAGGGGAAAAA 1822
 Db 142 GAAGAAGAAGATGATGATGAAGATGAAGAAGAA 110

RESULT 10
 BHI60272
 LOCUS ENTQV49TR Entamoeba histolytica Sheared DNA linear GSS 24-SEP-2001
 DEFINITION genomic, genomic survey sequence.
 ACCESSION BHI60272
 VERSION BHI60272.1 GI:15733710
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 931)
 AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HM1:IMSS sheared DNA library (2001)
 COMMENT Unpublished (2001)
 CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjl@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 8
 High quality sequence stop: 677.
 Location/Qualifiers
 FEATURES
 source
 1..931
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: pHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

Query Match 2.5%; Score 59.2; DB 28; Length 931;
 Best Local Similarity 43.3%; Pred. No. 0.0038;
 Matches 277; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

QY 1583 GATGGTTACATTTTGTGATGACATGATATATCATGATGATGAAGGATGCGATATGTAACG 1642
 DB 17 GATGATGATGAAGAAGAAGATGATGATGAAGAAGACGATGAAGACGACGAA 76

QY 1643 CCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAGCCCTTCTGTATAAGGAAAAAGTT 1702
 DB 77 TATGATTTAGAGATGATGATGATGATGAAGAAGATGATGATGAAGAAGAAGAT 136

QY 1703 GCAGTCAAGCCTATACTAAGAAAAGGTATCCTACCTCCATCTCCAGACGCGATGTT 1762
 DB 137 GATGATGAAGATGATGATGAAGAAGAAGATGAAGAAGATGATGAAGAAGAT 196

QY 1763 AAAGCAATCAACTGGAGATAGTGCACAGCTATTATCAATCGTGTGAAAGGGAAGAA 1822
 DB 197 GATGATGAAGAAGAAGATGATGATGAAGAAGATGAAGCGATGAAGACGACGATAT 256

QY 1823 CGAATTCACCTCGTTCCGATTTCCATATATGTTGAGCATACAGTTGAGGTTTAAACGGT 1882
 DB 257 GAATTAGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGAT 316

QY 1883 AATTGGATTATCTCATAGAGATCAATACCATATATTAATTGCTTGGTTTGTGATGAT 1942
 DB 317 GAAGATGATGATGAAGAAGAAGATGAAGATGATGATGAAGAAGAAGATGAAGATGAT 376

QY 1943 CACACATCAAGCTCCAAATGGCTATACCTTGGAAAGATTTGTTTGCACCGATTAAAGTAC 2002
 DB 377 GAAGAAGAGATGATGATGAAGAAGAAGATGAAGAAGAGATGATGAAGAGATGAT 436

QY 2003 TACGTAGAACACCCGTGACGAACGTCACATCTTAATGATGGGCGCAATGCCAGTGAG 2062

Db 437 GAAGAAGAAGATGATGATGAAGACGACGAATATGAATTAGAAGATGAAGAAGAT 496
 QY 2063 CATGTCTTAGGCAAGAAAGACACACAGTGAAGATCCAAATTAAGAACTTCAAAGCGGTAA 2122
 Db 497 GATGAAGAAGATGATGAAGAAGAAGATGATGAAGAAGAAGATGATGATGATGATGAT 556
 QY 2123 GAGCCAGTAGAGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTA 2182
 Db 557 GAAGAAGATGATGATGAAGATGATGAAGAAGATGAAGATGATGAAGAAGATGAAGATGAT 616
 QY 2183 GAAGCCCAACTCAAAAGCAGAAAGTTTCTGCTTGGCGAAG 2222
 Db 617 GAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAG 656

RESULT 11
 CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL063921 GI:4941778
 VERSION AL063921.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR08K10"
 /clone_lib="RPCI-98"
 /note="end : TET3"

ORIGIN
 Query Match 2.5%; Score 58.6; DB 29; Length 1101;
 Best Local Similarity 16.2%; Pred. No. 0.0056;
 Matches 97; Conservative 252; Mismatches 251; Indels 0; Gaps 0;

QY 1355 AAAGCCTTGTTGNAATAAGGGTCGTAATCTCTGATTTCCAGCCTTAGACAAATTATTA 1414
 Db 1098 RRGDDTDRKDDDDWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKWTKW 1039
 QY 1415 GAAAGCCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1474
 Db 1038 WWWATWTDWTKWWWATAAKTDTATWTTAWRADWAGRGAGKRGKRDADTDAGAGR 979

Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 100
 High quality sequence stop: 872.
 Location/Qualifiers
 1. 905
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IM5"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Query Match 2.4%; Score 56.8; DB 28; Length 905;
 Best Local Similarity 44.8%; Pred. No. 0.014;
 Matches 269; Conservative 0; Mismatches 323; Indels 9; Gaps 1;
 1139 GTACGAAAGTTGGGGAAGGATATGTTTCAAGAAAGGCGCATCTCTCGTTATGCTTT 1198
 Db 780 GAATTAGAGAGATGATGATGATGAATTTGAATTAGAGAGATGATGATGATGAT 721
 Qy 1199 GCGAAGATTTACCATCTGAAACCTTTAAATCTTGAAGCAAGTTATCAAAACAGAG 1258
 Db 720 GATCAAGACGGAAGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGAT 661
 Qy 1259 AGTGTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGATT 1318
 Db 660 GATGATGAATTTGAATTAGAGATGAAGAGATGAAGATGATGATGATGATGATGAT 601
 Qy 1319 TATGATAAGCATATAATCTGTTAACTGAGGCTCATATATTTGGCATTCCTAGCACCATTACC 1369
 Db 600 GATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 541
 Qy 1370 AATAAGGTCCTAATCTGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGAT 1429
 Db 540 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 Qy 1430 GAATCGACTAATAAGAAATTTGGTATGATGATTTATTTGGCATTCCTAGCACCATTACC 1489
 Db 480 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 Qy 1490 CATTCAGAGGCTTTGGCAACCAAAATTTCTCAAAATTTAGTATGATGATGATGATGAT 1549
 Db 420 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 Qy 1550 ATTGCTCAATTAGCTGATAGTATACACGTCAGATGATGATGATGATGATGATGATGAT 1609
 Db 360 GATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 Qy 1610 ATAATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
 Db 300 TTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 Qy 1670 GGAAGAATAGCTTTCTGTAAGGAAAGATTTGAGCTTCAAGCTTCAAGCTTCAAGGAAA 1729
 Db 240 GAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
 Qy 1730 G 1730
 Db 180 G 180

RESULT 14

BM165350
 LOCUS
 DEFINITION
 EST567873 PyBS Plasmodium yoelii yoelii cDNA clone PYCML60 5' end,
 mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 645)
 Carlson,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Garucci,D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 CONTACT: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.
 Location/Qualifiers
 1. 645
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17X1"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCML60"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

FEATURES

source

ORIGIN

Query Match 2.4%; Score 56.6; DB 12; Length 645;
 Best Local Similarity 47.4%; Pred. No. 0.014;
 Matches 170; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
 107 GATGAGGTTACCAAGCGTGAAGGAATCAATCTCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 Db 120 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179
 Qy 167 CAAGCTTATGCTTTCATGCGGACCACTATCATTTATTCATGTTAAAGTTCTTAT 226
 Db 180 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239
 Qy 227 GACGCTATCATGATGAGAAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAG 286

Db 538 GAGGATGAGATGATGAAGAAGATGACGATGATTCTTAAAAAAGAGGTGGATCAAAATGC 596

RESULT 18
LOCUS BM160252
DEFINITION EST562775 PyBS Plasmodium yoelii cDNA clone PYCJ090 5' end,
mRNA sequence.
ACCESSION BM160252
VERSION BM160252.1 GI:17305933
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii
ORGANISM Plasmodium yoelii yoelii
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 671)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES
source
1..671
/organism="Plasmodium yoelii yoelii"
/mol_type="mRNA"
/strain="17XL"
/sub_species="yoelii"
/db_xref="taxon:73239"
/clone="PYCJ090"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/clone_lib="PyBS"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase, and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

ORIGIN
Query Match 2.3%; Score 55; DB 12; Length 671;
Best Local Similarity 47.1%; Pred. No. 0.035;
Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGTTAGCAAGGTTGAAGATCATGCTGACCAATCGTCATCAAGTAAACAGAC 166
Db |||||
194 GATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253
QY 167 CAAGGCTATGTCACATCGCGACCACTATCATTTATTCAATGGTAGGTTCTTAT 226
Db |||||
254 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 313
QY 227 GACGCTATCATCAGTGAAGATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAG 286

Db 314 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 373
QY 287 GATATTGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGATAAATACTATGTT 346
Db 374 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 433
QY 347 TACCTTAAGGATGCTGCCACGCGGATPAACGTCCGTACAAAAGAGGAAATCAATCGCAA 406
Db 434 GACGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 493
QY 407 AAACAAGAGCATAGTCAACATCGTGAAGTGGAACTCCRAAGAACCGATGCTGTTGCC 465
Db 494 GAGGATGAAGATGATGAAGAAGATGACGATGATTCTTAAAAAAGAGGTGGATCAAAATGC 552

RESULT 19
LOCUS BM168242
DEFINITION EST570765 PyBS Plasmodium yoelii cDNA clone PYCJ060 5' end,
mRNA sequence.
ACCESSION BM168242
VERSION BM168242.1 GI:17301474
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii
ORGANISM Plasmodium yoelii yoelii
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 694)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES
source
1..694
/organism="Plasmodium yoelii yoelii"
/mol_type="mRNA"
/strain="17XL"
/sub_species="yoelii"
/db_xref="taxon:73239"
/clone="PYCJ060"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/clone_lib="PyBS"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase, and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 694;
 Best Local Similarity 47.1%; Pred. No. 0.035;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 Db 192 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGACGATGAC 251
 |||||

QY 167 CAGGCTATGTCACCTTCCATGCGGACACATATATTTACATGTTAAGTTTCCTTAT 226
 |||||
 Db 252 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
 |||||

QY 227 GAGCTATCATGCTGAGGAATTAATCATGAAACATCAAACTATAAGCTAAAAGATGAG 286
 |||||
 Db 312 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
 |||||

QY 287 GATATTGTTAATGAGGTCAGGGTGGATATGTTATCAAGGTPAGATGAGAAATATCTATGTT 346
 |||||
 Db 372 GACGAT 431
 |||||

QY 347 TACCTTAAGGATGCTCCAGCGGATTAACGTCGTCACAAAGAGGAATCAATCGACAA 406
 |||||
 Db 432 GACGAT 491
 |||||

QY 407 AAACAAGAGCATGCTCAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGTTGC 465
 |||||
 Db 492 GAGGATGAAGATGATGAAGAGATGACGATGATCTTAAAGAAAGAGTGGATCAATGC 550
 |||||

RESULT 20
 BM160500 717 bp mRNA linear EST 04-DEC-2001
 LOCUS EST563023 PyBS Plasmodium yoelii yoe111 cDNA clone PYCJS13 5' end,
 DEFINITION mRNA sequence.
 ACCESSION BM160500
 VERSION BM160500.1 GI:17306181
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoe111
 ORGANISM Plasmodium yoelii yoe111
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 717)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 source
 1..717
 /organism="Plasmodium yoelii yoe111"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoe111"
 /db_xref="taxon:73239"
 /clone="PYCJS13"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography."

First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pAD-GAL4) was
 excised from the HybriZAP vector and plasmid DNA
 isolated."

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 717;
 Best Local Similarity 47.1%; Pred. No. 0.036;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 Db 205 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGACGATGAC 264
 |||||

QY 167 CAAGGCTATGTCACATTCACATGCGCACCTATCATTTATTAATGTTAAGTTTCCTTAT 226
 |||||
 Db 265 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
 |||||

QY 227 GACGCTATCATGCTGAGGAATTAATCATGAAAGATCCAAACTATAAGCTAAAAGATGAG 286
 |||||
 Db 325 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
 |||||

QY 287 GATATTGTTAATGAGGTCAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTT 346
 |||||
 Db 385 GACGAT 444
 |||||

QY 347 TACCTTAAGGATGCTCCAGCGGATTAACGTCGTCACAAAGAGGAATCAATCGACAA 406
 |||||
 Db 445 GACGAT 504
 |||||

QY 407 AAACAAGAGCATGCTCAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGTTGC 465
 |||||
 Db 505 GAGGATGAAGATGATGAAGAGATGACGATGATCTTAAAGAAAGAGTGGATCAATGC 563
 |||||

RESULT 21
 BM162492 756 bp mRNA linear EST 04-DEC-2001
 LOCUS EST565015 PyBS Plasmodium yoelii yoe111 cDNA clone PYCKT28 5' end,
 DEFINITION mRNA sequence.
 ACCESSION BM162492
 VERSION BM162492.1 GI:17308173
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoe111
 ORGANISM Plasmodium yoelii yoe111
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 756)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 Location/Qualifiers
 1..756
 /organism="Plasmodium yoelii yoe111"
 /mol_type="mRNA"

/strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCK28"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was mRNA isolated using the guanidinium isothiocyanate method, and First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 756;
 Best Local Similarity 47.1%; Pred. No. 0.036;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 Db 113 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 172
 |||||
 QY 167 CAAGCTATGTCACTTCACATGCGACCACTATCATTTTCAATGATGATGATGATGATGAT 226
 |||||
 Db 173 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 232
 |||||
 QY 227 GACGCTATCATCATGATGAAGATTTACTCATGAAGATCCAAACTATTAAGTAAAGATGAG 286
 |||||
 Db 233 GAAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292
 |||||
 QY 287 GATATTGTTAATGAGTCAAGGTTGGATATGTTTATCAAGGTAGATGCAAAATACTATGTT 346
 |||||
 Db 293 GACGATGATGATGATGACGACGATGAAGATTTTGAAGATATGATGATGATGATGATGAT 352
 |||||
 QY 347 TACCTTAAGGATGCTCCACCGGTAACGTCCTGACAAAAGAGGAATCAATCGACAA 406
 |||||
 Db 353 GACGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 412
 |||||
 QY 407 AAACAGAGCATAGTCAACATCGTGAAGGTGGAACCTCCAAAGAACGATGCTGCTGTCG 465
 |||||
 Db 413 GAGGATGAAGATGATGAAGAGATGACGATGATTTCTAAAGAAAGAGGTGGATCAATGC 471
 |||||

RESULT 22
 BM159906
 LOCUS
 DEFINITION EST562429 PyBS Plasmodium yoelii yoe11i cDNA clone PYCK35 5' end, mRNA sequence.
 ACCESSION BM159906
 VERSION BM159906.1 GI:17305587
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoe11i
 ORGANISM Plasmodium yoelii yoe11i
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 794)
 Carleton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Garucci, D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carleton
 Parasite Genomics Group

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source 1..794
 Location/Qualifiers
 /organism="Plasmodium yoelii yoe11i"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCK35"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was mRNA isolated using the guanidinium isothiocyanate method, and First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 794;
 Best Local Similarity 47.1%; Pred. No. 0.037;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 Db 205 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 264
 |||||
 QY 167 CAAGGCTATGTCACATCGCGACCACTATCATTTTCAATGATGATGATGATGATGATGATGAT 226
 |||||
 Db 265 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
 |||||
 QY 227 GACGCTATCATCAGTGAAGAAATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAG 286
 |||||
 Db 325 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
 |||||
 QY 287 GATATTGTTAATGAGTCAAGGTTGGATATGTTTATCAAGGTAGATGCAAAATACTATGTT 346
 |||||
 Db 385 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
 |||||
 QY 347 TACCTTAAGGATGCTCCACCGGATTAACGTCCTGACAAAAGAGGAATCAATCGACAA 406
 |||||
 Db 445 GACGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
 |||||
 QY 407 AAACAGAGCATAGTCAACATCGTGAAGGTGGAACCTCCAAAGAAACGATGCTGCTGTCG 465
 |||||
 Db 505 GAGGATGAAGATGATGAAGAGATGACGATGATTTCTAAAGAAAGAGGTGGATCAATGC 563
 |||||

RESULT 23
 AZ528485
 LOCUS
 DEFINITION ENT528485 Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
 795 bp DNA linear GSS 03-NOV-2000

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 15:35:29 ; Search time 170. Seconds
(without alignments)
7798.692 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTGGGACTGT.....TAAGTAAGGAAAAATAAC 2389

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/ECTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2388	100.0	2389	3	US-08-961-083-55
2	2388	100.0	2389	4	US-09-536-784-55
3	2388	100.0	2451	4	US-09-468-656A-9
4	2388	100.0	8195	4	US-08-961-527-94
5	991	41.5	2531	4	US-09-468-656A-11
6	990	41.4	2531	4	US-09-468-656A-5
7	987.6	41.3	2290	3	US-08-961-083-65
8	987.6	41.3	2290	4	US-09-536-784-65
9	653.8	27.4	2359	4	US-08-961-527-243
10	385.4	16.1	1342	3	US-08-961-083-181
11	385.4	16.1	1342	4	US-09-536-784-181
12	385.4	16.1	1455	4	US-09-468-656A-7
13	385.4	16.1	6867	4	US-08-961-527-192
14	381.4	16.0	973	4	US-08-961-527-358
15	243.4	10.2	1684	4	US-08-961-527-255
16	109	4.6	841	2	US-08-743-637B-34
17	109	4.6	841	3	US-08-526-840B-34
18	57.4	2.4	7218	1	US-08-232-463-14
19	51.6	2.2	11091	4	US-09-134-001C-2243
20	45.2	1.9	2277	1	US-08-676-967-2
21	45.2	1.9	2277	1	US-08-676-974-2
22	45.2	1.9	2277	2	US-09-098-487-2
23	44	1.8	832	4	US-09-621-976-2813
24	42.4	1.8	3945	4	US-09-200-650E-6
25	42.4	1.8	4358	4	US-08-956-171E-454
26	42	1.8	6948	4	US-09-543-681A-1262
27	41.4	1.7	5361	3	US-08-973-462-2

ALIGNMENTS

28	41.4	1.7	6152	3	US-08-973-462-1	Sequence 1, Appli
29	41.2	1.7	478	4	US-09-621-976-10407	Sequence 10407, A
30	41.2	1.7	696	3	US-09-461-697-193	Sequence 193, App
31	41.2	1.7	699	3	US-09-461-697-191	Sequence 191, App
32	41.2	1.7	717	3	US-09-461-697-189	Sequence 189, App
33	41.2	1.7	774	3	US-09-461-697-187	Sequence 187, App
34	41.2	1.7	819	3	US-09-461-697-185	Sequence 185, App
35	41.2	1.7	1280	3	US-09-220-133-155	Sequence 155, App
36	41.2	1.7	1669	3	US-09-461-697-184	Sequence 184, App
37	41	1.7	5340	4	US-09-627-122-21	Sequence 21, Appl
38	41	1.7	14066	4	US-09-601-198-56	Sequence 56, Appl
39	40.6	1.7	1704	4	US-09-134-001C-925	Sequence 925, App
40	40.4	1.7	1332	4	US-09-543-681A-819	Sequence 819, App
41	40.2	1.7	792	4	US-09-328-352-2472	Sequence 2472, App
42	40	1.7	30549	4	US-09-134-001C-322	Sequence 322, App
43	39.8	1.7	297	4	US-09-134-001C-1027	Sequence 1027, App
44	39.8	1.7	1716	4	US-09-134-001C-1028	Sequence 1028, App
45	39.6	1.7	1308	4	US-09-601-198-3	Sequence 3, Appli
46	39.6	1.7	18355	4	US-08-956-171E-67	Sequence 67, Appli
47	39.4	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
48	39.2	1.6	929	4	US-09-671-317-14	Sequence 14, Appl
49	39	1.6	1839	4	US-09-134-001C-1191	Sequence 1191, App
50	38.8	1.6	1856	2	US-08-360-606B-29	Sequence 29, Appl
51	38.8	1.6	2150	2	US-08-861-464-13	Sequence 13, Appl
52	38.8	1.6	2150	2	US-08-396-001-13	Sequence 13, Appl
53	38.8	1.6	2150	3	US-09-323-433A-13	Sequence 13, Appl
54	38.6	1.6	6744	1	US-08-119-125A-2	Sequence 2, Appli
55	38.6	1.6	8920	2	US-08-446-855A-1	Sequence 1, Appli
56	38.6	1.6	8920	3	US-09-150-741-1	Sequence 1, Appli
57	38.4	1.6	5319	1	US-08-169-927-1	Sequence 1, Appli
58	38.4	1.6	6846	4	US-08-961-527-198	Sequence 198, App
59	38.4	1.6	580073	4	US-08-545-528D-1	Sequence 1, Appli
60	38.2	1.6	1866	4	US-09-601-198-153	Sequence 153, App
61	38.2	1.6	640681	4	US-09-790-988-1	Sequence 1, Appli
62	38	1.6	1001	4	US-09-671-317-439	Sequence 439, App
63	38	1.6	3280	1	US-08-259-000-4	Sequence 4, Appli
64	38	1.6	3280	1	US-08-729-767-6	Sequence 6, Appli
65	38	1.6	8536	4	US-08-956-171E-278	Sequence 278, App
66	37.8	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
67	37.6	1.6	438	4	US-09-134-001C-1064	Sequence 1064, App
68	37.6	1.6	2000	4	US-09-632-313A-1	Sequence 1, Appli
69	37.6	1.6	2000	4	US-09-921-942B-6	Sequence 6, Appli
70	37.6	1.6	2787	4	US-09-134-000C-3185	Sequence 3185, App
71	37.4	1.6	3294	3	US-08-923-992A-7	Sequence 7, Appli
72	37.4	1.6	26385	4	US-08-961-527-3	Sequence 3, Appli
73	37.2	1.6	462	4	US-09-328-352-2077	Sequence 2077, App
74	37.2	1.6	1079	1	US-07-781-355-1	Sequence 1, Appli
75	37.2	1.6	2397	4	US-09-134-001C-2080	Sequence 2080, App
76	37.2	1.6	1830121	4	US-09-557-884-1	Sequence 1, Appli
77	37.2	1.6	1830121	4	US-09-643-990A-1	Sequence 1, Appli
78	37	1.5	1103	3	US-09-123-400B-20	Sequence 20, Appli
79	37	1.5	1167	1	US-08-592-658-1	Sequence 1, Appli
80	37	1.5	1485	4	US-09-543-681A-601	Sequence 601, App
81	37	1.5	1893	4	US-09-328-352-112	Sequence 112, App
82	37	1.5	9310	4	US-08-956-171E-316	Sequence 316, App
83	36.6	1.5	282	3	US-09-461-697-205	Sequence 205, App
84	36.6	1.5	306	3	US-09-461-697-203	Sequence 203, App
85	36.6	1.5	807	4	US-09-328-352-2911	Sequence 2911, App
86	36.6	1.5	1065	4	US-09-107-532A-3579	Sequence 3579, App
87	36.6	1.5	1797	4	US-09-134-000C-1552	Sequence 1552, App
88	36.6	1.5	2145	4	US-09-059-584-48	Sequence 48, Appli
89	36.6	1.5	2287	4	US-09-059-584-47	Sequence 47, Appli
90	36.6	1.5	3312	4	US-09-601-198-56	Sequence 58, Appli

RESULT 1

US-08-961-083-55

; Sequence 55, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,093
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-55

Query Match 100.0%; Score 2389; DB 3; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGTTAAGGAAATAATCGTGTTCCTA	60
Db	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGTTAAGGAAATAATCGTGTTCCTA	60
Qy	61	TATAGATGGAACCAAGCGAGCGCAAAACCGGAAATTTGACTCCTGATGAGGTTAGCAA	120
Db	61	TATAGATGGAACCAAGCGAGCGCAAAACCGGAAATTTGACTCCTGATGAGGTTAGCAA	120
Qy	121	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC	180
Db	121	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC	180
Qy	181	TTCCATGCGGCAACCATATCAATTAATCAATGTTAAGGTTCCCTATGACGCTATCATCAG	240
Db	181	TTCCATGCGGCAACCATATCAATTAATCAATGTTAAGGTTCCCTATGACGCTATCATCAG	240
Qy	241	TGAAGAATTACTCATGAAAGATCCAACTAAGCTAAAGATGAGGATATTTGTTAATGA	300
Db	241	TGAAGAATTACTCATGAAAGATCCAACTAAGCTAAAGATGAGGATATTTGTTAATGA	300
Qy	301	GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Db	301	GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Qy	361	TGCCCGACGGGATTAACGTCGCTACAAAGAGGAATCAATGACAAACCAAGAGCATAG	420
Db	361	TGCCCGACGGGATTAACGTCGCTACAAAGAGGAATCAATGACAAACCAAGAGCATAG	420
Qy	421	TCAACATCGTGAAGTGAATCCAAAGAAACGATGGTGTGCTTGGCTTGGCACGTTCCGA	480
Db	421	TCAACATCGTGAAGTGAATCCAAAGAAACGATGGTGTGCTTGGCTTGGCACGTTCCGA	480

Qy	481	AGGACCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGTAC	540
Db	481	AGGACCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGTAC	540
Qy	541	TGCTGATGCTTATATGTTTCTTCATCGAGATCATACCATATCATTCCTTAAGATGATT	600
Db	541	TGCTGATGCTTATATGTTTCTTCATCGAGATCATATACCATATCATTCCTTAAGATGATT	600
Qy	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTTGGTCGAGGAAATCTGTCAA	660
Db	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTTGGTCGAGGAAATCTGTCAA	660
Qy	661	TTCAAGAACCTATCGCCGACAAATAGCGATAACACTTTCAAGAACCAACTGGGTACCTTC	720
Db	661	TTCAAGAACCTATCGCCGACAAATAGCGATAACACTTTCAAGAACCAACTGGGTACCTTC	720
Qy	721	TGTAAGCAATCCAGGAACCTAACAATTAACAAGCAACCAAGCAACCACTAACAGTCA	780
Db	721	TGTAAGCAATCCAGGAACCTAACAATTAACAAGCAACCAAGCAACCACTAACAGTCA	780
Qy	781	AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTTACAACTGCCTTTGAG	840
Db	781	AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTTACAACTGCCTTTGAG	840
Qy	841	TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCAAC	900
Db	841	TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCAAC	900
Qy	901	AGCTAGAGGTTGCGAGTCCACGAGATCATACCACTTCATCCCTTACTCTCAAAAT	960
Db	901	AGCTAGAGGTTGCGAGTCCACGAGATCATACCACTTCATCCCTTACTCTCAAAAT	960
Qy	961	GTCTGAATTTGGAAGAACGAATCGTGTATTTTCCCTTTCGTTTATCGTTTCAAAACCATTTG	1020
Db	961	GTCTGAATTTGGAAGAACGAATCGTGTATTTTCCCTTTCGTTTATCGTTTCAAAACCATTTG	1020
Qy	1021	GGTACCAGATTCGAAGCCAGAACCAACCAAGTCCCAACCGGAACTCCGAACTAGTCCAGG	1080
Db	1021	GGTACCAGATTCGAAGCCAGAACCAACCAAGTCCCAACCGGAACTCCGAACTAGTCCAGG	1080
Qy	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTCAAGTGGT	1140
Db	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTCAAGTGGT	1140
Qy	1141	ACGAAAAGTTGGGAGAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGCTTTGC	1200
Db	1141	ACGAAAAGTTGGGAGAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGCTTTGC	1200
Qy	1201	GAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG	1260
Db	1201	GAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG	1260
Qy	1261	TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA	1320
Db	1261	TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA	1320
Qy	1321	TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGGTCG	1380
Db	1321	TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGGTCG	1380
Qy	1381	TAATCTGATTTCAAGCCTTAGCAAAATTTATAGACGCTTGAATGATGATCGACTAA	1440
Db	1381	TAATCTGATTTCAAGCCTTAGCAAAATTTATAGACGCTTGAATGATGATCGACTAA	1440
Qy	1441	TAAAGAAAATTTGGTAGATGATTTATGGCAATCTTAGCACAATTAACCATCCAGAGCG	1500
Db	1441	TAAAGAAAATTTGGTAGATGATTTATGGCAATCTTAGCACAATTAACCATCCAGAGCG	1500
Qy	1501	ACTTGGCAACCAAAATCTCAAAATGAGATATCTGAAGACGAAGTTGTTGCTCAAT	1560
Db	1501	ACTTGGCAACCAAAATCTCAAAATGAGATATCTGAAGACGAAGTTGTTGCTCAAT	1560
Qy	1561	AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGGATGAACATGATATAATCAGTGA	1620


```

661 QY TTCAAGAACTATCGCGCAAAATAGCGATACACCTTCAAGAACAACTGGGTACCTTC 720
661 Db TTCAAGAACTATCGCGCAAAATAGCGATACACCTTCAAGAACAACTGGGTACCTTC 720
721 QY TGTAAAGCAATCCAGGAACTACAATACTACACAGCAACCAAGCAACAACTAACAGTCA 780
721 Db TGTAAAGCAATCCAGGAACTACAATACTACACAGCAACCAAGCAACAACTAACAGTCA 780
781 QY AGCAAGTCAAAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTAACAACCTGCTTTGAG 840
781 Db AGCAAGTCAAAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTAACAACCTGCTTTGAG 840
841 QY TCAACGACATGTAGATCTGATGGCCTGCTTTGATCCAGCAACAATGACAACTGCAAC 900
841 Db TCAACGACATGTAGATCTGATGGCCTGCTTTGATCCAGCAACAATGACAACTGCAAC 900
901 QY AGCTAGAGGTGTTGCAAGTGCACACGAGAGATCAATACCACTTCATCCCTTACTCTCAAAAT 960
901 Db AGCTAGAGGTGTTGCAAGTGCACACGAGAGATCAATACCACTTCATCCCTTACTCTCAAAAT 960
961 QY GTCTGAATTTGGAAGAACGAATCGCTCGTATATTTCCCTTCCTGTTATCGTTCAAAACCAATTG 1020
961 Db GTCTGAATTTGGAAGAACGAATCGCTCGTATATTTCCCTTCCTGTTATCGTTCAAAACCAATTG 1020
1021 QY GGTACCAAGATTCAGGCGCAGAACCAAGTCCACACCGACTCCCGAACCTAGTCCAGG 1080
1021 Db GGTACCAAGATTCAGGCGCAGAACCAAGTCCACACCGACTCCCGAACCTAGTCCAGG 1080
1081 QY CCGCAACCTGCACCAAACTCTTAAATAGACTCAAAATCTCTTTGTTAGTACGCTGGT 1140
1081 Db CCGCAACCTGCACCAAACTCTTAAATAGACTCAAAATCTCTTTGTTAGTACGCTGGT 1140
1141 QY ACGAAAAGTTGGGAAGAGATATGATTCGAAGAAAGGGGCAATCTCTGTTATGCTTTTGC 1200
1141 Db ACGAAAAGTTGGGAAGAGATATGATTCGAAGAAAGGGGCAATCTCTGTTATGCTTTTGC 1200
1201 QY GAAAGATTTACCACTCTGAACTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
1201 Db GAAAGATTTACCACTCTGAACTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
1261 QY TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCCTGCTGACCAAGAAATTTA 1320
1261 Db TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCCTGCTGACCAAGAAATTTA 1320
1321 QY TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTGNAATAAAGGTCG 1380
1321 Db TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTGNAATAAAGGTCG 1380
1381 QY TAATTCGTATTTCCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1381 Db TAATTCGTATTTCCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1441 QY TAAAGAAAAATGTTAGATGATTTATTGGCAATTCCTAGCACCAATACCCATCCAGAGCG 1500
1441 Db TAAAGAAAAATGTTAGATGATTTATTGGCAATTCCTAGCACCAATACCCATCCAGAGCG 1500
1501 QY ACTTGGCAAAACCAATCTTCAAAATGAGTACTGAGACGAAGTTGTTATGCTCAATT 1560
1501 Db ACTTGGCAAAACCAATCTTCAAAATGAGTACTGAGACGAAGTTGTTATGCTCAATT 1560
1561 QY AGCTGATAAGTATCAACGTCAGATGTTACATTTTGTGAAACATGATATAATCACTGA 1620
1561 Db AGCTGATAAGTATCAACGTCAGATGTTACATTTTGTGAAACATGATATAATCACTGA 1620
1621 QY TGAAGGAGATGCATATGTAAAGCCTCATATATGGGCCATAGTCACTGGATGGAAAGATAG 1680
1621 Db TGAAGGAGATGCATATGTAAAGCCTCATATATGGGCCATAGTCACTGGATGGAAAGATAG 1680
1681 QY CCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTTACTAAGAAAGGATTCCTACC 1740
1681 Db CCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTTACTAAGAAAGGATTCCTACC 1740

```

```

1741 QY TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGACGACTATTTA 1800
1741 Db TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGACGACTATTTA 1800
1801 QY CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
1801 Db CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
1861 QY TACAGTTGAGGTTAAAGCGGTAATTTGATTTATCTCTATAAGGATCATTTACCATATAT 1920
1861 Db TACAGTTGAGGTTAAAGCGGTAATTTGATTTATCTCTATAAGGATCATTTACCATATAT 1920
1921 QY TAAATTTGCTTTGCTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
1921 Db TAAATTTGCTTTGCTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
1981 QY TTTGTTTGGGACGATTAAGTACTACGTAGAACAACCTGACGAACGCTCCACATTTCTAATGA 2040
1981 Db TTTGTTTGGGACGATTAAGTACTACGTAGAACAACCTGACGAACGCTCCACATTTCTAATGA 2040
2041 QY TGGATCGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
2041 Db TGGATCGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
2101 QY TAAGAACTTCAAGCGGATGAAGCGGATGAAGCGGATGAAGGAAACACCTGCTGAGCAGAGTCCC 2160
2101 Db TAAGAACTTCAAGCGGATGAAGCGGATGAAGCGGATGAAGGAAACACCTGCTGAGCAGAGTCCC 2160
2161 QY TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTTGGGAA 2220
2161 Db TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTTGGGAA 2220
2221 QY AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
2221 Db AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
2281 QY TTTGACTCTTTCAAAATTTAGTAACTAATAGTATCATGGCAGAGAGAGAAAAATTTACTTGC 2340
2281 Db TTTGACTCTTTCAAAATTTAGTAACTAATAGTATCATGGCAGAGAGAGAAAAATTTACTTGC 2340
2341 QY GTTGTAAAGGAAGTAACTCTTCACTGTAAAGTAAAGGAAAAAATAAAC 2389
2341 Db GTTGTAAAGGAAGTAACTCTTCACTGTAAAGTAAAGGAAAAAATAAAC 2389

```

```

RESULT 3
US-09-468-656A-9
; Sequence 9, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2451)
; OTHER INFORMATION: n = a, c, t or g
US-09-468-656A-9
Query Match 100.0%; Score 2388; DB 4; Length 2451;

```

Best Local Similarity 100.0%; Pred. No. 0;		Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA	60
Db	60	TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA	119
QY	61	TATAGATGGAAGAACAAAGCGACGCAAAAGGCGGAGAAATTTGACCTCTGATGAGGTAGCAA	120
Db	120	TATAGATGGAAGAACAAAGCGACGCAAAAGGCGGAGAAATTTGACCTCTGATGAGGTAGCAA	179
QY	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGCAAGCTATGTCA	180
Db	180	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGCAAGCTATGTCA	239
QY	181	TTACATGCGGACCACTATCATTTATTAATGATGTAAGTTCCTTATGACGCTATCATCAG	240
Db	240	TTACATGCGGACCACTATCATTTATTAATGATGTAAGTTCCTTATGACGCTATCATCAG	299
QY	241	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGA	300
Db	300	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGA	359
QY	301	GCTCAAGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Db	360	GCTCAAGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	419
QY	361	TGCCACGCGGATAACGTCGTCACAAAGAGGAAATCAATCGCAAAAGCAAGAGCATAG	420
Db	420	TGCCACGCGGATAACGTCGTCACAAAGAGGAAATCAATCGCAAAAGCAAGAGCATAG	479
QY	421	TCAACATCTGAGGTGGAACTCCAAAGAAACGATGGTGTGTTGCTTCGCGAGTTCGCA	480
Db	480	TCAACATCTGAGGTGGAACTCCAAAGAAACGATGGTGTGTTGCTTCGCGAGTTCGCA	539
QY	481	AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC	540
Db	540	AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC	599
QY	541	TGCTGATGCTTATATCGTTCCTCATGAGATCATTTACCATTTACATTCCTTAAGATGAT	600
Db	600	TGCTGATGCTTATATCGTTCCTCATGAGATCATTTACCATTTACATTCCTTAAGATGAT	659
QY	601	ATCAGCTAGCGAGTTGGCTGCTGCAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA	660
Db	660	ATCAGCTAGCGAGTTGGCTGCTGCAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA	719
QY	661	TTCAAGAACCTATGCGCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC	720
Db	720	TTCAAGAACCTATGCGCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC	779
QY	721	TGTAAGCAATCCAGGAATACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	780
Db	780	TGTAAGCAATCCAGGAATACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	839
QY	781	AGCAAGTCAAAAGTAATGACATTTAGTAGTCTCTTTGAAACAGCTCTCAAACTCCCTTTG	840
Db	840	AGCAAGTCAAAAGTAATGACATTTAGTAGTCTCTTTGAAACAGCTCTCAAACTCCCTTTG	899
QY	841	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAAATCACAAGTCGA	900
Db	900	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAAATCACAAGTCGA	959
QY	901	AGCTAGAGTGTTCAGTGCACAGGAGATCATTTACGATTCCTTCTTCTCAAT	960
Db	960	AGCTAGAGTGTTCAGTGCACAGGAGATCATTTACGATTCCTTCTTCTCAAT	1019
QY	961	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTTCTGTTATCGTTCAAACTTG	1020
Db	1020	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTTCTGTTATCGTTCAAACTTG	1079
QY	1021	GGTACAGATTCAGGCGGAGCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG	1080

Db	1080	GGTACAGATTCAGGCGGAGCAACCAAGTCCACAAACCGACTCCGGAACTAGTCCAGG	1139
QY	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTTCTTTTGGTTAGTCCAGCTGGT	1140
Db	1140	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTTCTTTTGGTTAGTCCAGCTGGT	1199
QY	1141	ACGAAAGTTGGGGAAGGATATGTTTTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGC	1200
Db	1200	ACGAAAGTTGGGGAAGGATATGTTTTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGC	1259
QY	1201	GAAGAGTTTACCATCTGAAACCTGTAATAATCTTGAAGCAAGTTATCAAAACAAGAG	1260
Db	1260	GAAGAGTTTACCATCTGAAACCTGTAATAATCTTGAAGCAAGTTATCAAAACAAGAG	1319
QY	1261	TGTTTCAACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAATTTA	1320
Db	1320	TGTTTCAACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAATTTA	1379
QY	1321	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG	1380
Db	1380	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG	1439
QY	1381	TAATTCCTGATTTTCAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATCAACTAA	1440
Db	1440	TAATTCCTGATTTTCAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATCAACTAA	1499
QY	1441	TAAAGAAAAATTTGGTAGATGATTTATTGGCAITTCCTAGCACCAATTAACCATCCAGCG	1500
Db	1500	TAAAGAAAAATTTGGTAGATGATTTATTGGCAITTCCTAGCACCAATTAACCATCCAGCG	1559
QY	1501	ACTTGGCAACCAATTTCTCAATTTGAGTATATCTGAAGACGAGTTGCTTATGCTCAAT	1560
Db	1560	ACTTGGCAACCAATTTCTCAATTTGAGTATATCTGAAGACGAGTTGCTTATGCTCAAT	1619
QY	1561	AGCTGATAGTATACAACTGAGTGTACATTTTGTGATGAACATGATATATCACTGCA	1620
Db	1620	AGCTGATAGTATACAACTGAGTGTACATTTTGTGATGAACATGATATATCACTGCA	1679
QY	1621	TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCTCATGGATTGGAAGAGATAG	1680
Db	1680	TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCTCATGGATTGGAAGAGATAG	1739
QY	1681	CTTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATATAAGAAAAAGGATATCTAC	1740
Db	1740	CTTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATATAAGAAAAAGGATATCTAC	1799
QY	1741	TCCATCTCCAGACGAGATGTTTAAAGCAAACTCAAACCTGAGATAGTGAAGCAAGGATAT	1800
Db	1800	TCCATCTCCAGACGAGATGTTTAAAGCAAACTCAAACCTGAGATAGTGAAGCAAGGATAT	1859
QY	1801	CAATCGTGTGAAAGGGGAAAAACGAATTCCTCGTTCCATTCATATATGTTGAGCA	1860
Db	1860	CAATCGTGTGAAAGGGGAAAAACGAATTCCTCGTTCCATTCATATATGTTGAGCA	1919
QY	1861	TACAGTTCAGGTTTAAAGCGGTAAATTTGATTTATTCCTATAAGGATCATTAATATAT	1920
Db	1920	TACAGTTCAGGTTTAAAGCGGTAAATTTGATTTATTCCTATAAGGATCATTAATATAT	1979
QY	1921	TAAATTTGCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGCTATACCTCGA	1980
Db	1980	TAAATTTGCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGCTATACCTCGA	2039
QY	1981	TTTGTTCGACGATTAAGTACTACGTAGAACACCTCGACGAACGTCCACATTTCTAATGA	2040
Db	2040	TTTGTTCGACGATTAAGTACTACGTAGAACACCTCGACGAACGTCCACATTTCTAATGA	2099
QY	2041	TGGATGGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAACACACAGTGAAGATCCAAA	2100
Db	2100	TGGATGGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAACACACAGTGAAGATCCAAA	2159
QY	2101	TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGAACACCTGCTGAGCGAAGTCCC	2160
Db	2160	TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGAACACCTGCTGAGCGAAGTCCC	2219

Qy	2161	TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAGAAGCAGAGAGTTTGGCTTGGAA	2220	1173	GGCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	3232
Db	2220	TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAGAAGCAGAGAGTTTGGCTTGGAA	2279	181	TTCAATATGGGACCACTATCAATTTATACATGTTAGAGTTCCCTTATGACGCTATCATCAG	240
Qy	2221	AGTAAAGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA	2280	3233	TTCAATATGGGACCACTATCAATTTATACATGTTAGAGTTCCCTTATGACGCTATCATCAG	3292
Db	2280	AGTAAAGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA	2339	241	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAAATGA	300
Qy	2281	TTTGACTCTTCAAATTTATGATTAACAATAGTATCATGCGAGAGCAGCAAAAATTTACTTGC	2340	3293	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAAATGA	3352
Db	2340	TTTGACTCTTCAAATTTATGATTAACAATAGTATCATGCGAGAGCAGCAAAAATTTACTTGC	2399	301	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGATGC	360
Qy	2341	GTGTTAAAGGAAGTAAATCTTCACTCTGTAAGTAAGGAAAAATAAAC	2389	3353	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGATGC	3412
Db	2400	GTGTTAAAGGAAGTAAATCTTCACTCTGTAAGTAAGGAAAAATAAAC	2448	361	TGCCCAGCGGATAACGTCGCTACAAAGAGGAATCAATCGCAAAAAAACAAGAGCATAG	420
RESULT 4						
US-08-961-527-94						
; Sequence 94, Application US/08961527						
; Patent No. 6420135						
GENERAL INFORMATION:						
; APPLICANT: Charles Kunsch						
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences						
; NUMBER OF SEQUENCES: 391						
; CORRESPONDENCE ADDRESS:						
; ADDRESSEE: Human Genome Sciences, Inc.						
; STREET: 9410 Key West Avenue						
; CITY: Rockville						
; STATE: Maryland						
; COUNTRY: USA						
; ZIP: 20850						
; COMPUTER READABLE FORM:						
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage						
; COMPUTER: HP Vectra 486/33						
; OPERATING SYSTEM: MSDOS version 6.2						
; SOFTWARE: ASCII Text						
; CURRENT APPLICATION DATA:						
; APPLICATION NUMBER: US/08/961,527						
; FILING DATE:						
; CLASSIFICATION: 424						
; PRIOR APPLICATION DATA:						
; APPLICATION NUMBER:						
; FILING DATE:						
; ATTORNEY/AGENT INFORMATION:						
; NAME: Brookes, A. Anders						
; REGISTRATION NUMBER: 36,373						
; REFERENCE/DOCKET NUMBER: PB340P1						
; TELECOMMUNICATION INFORMATION:						
; TELEPHONE: (301) 309-8504						
; TELEFAX: (301) 309-8512						
; INFORMATION FOR SEQ ID NO: 94:						
; SEQUENCE CHARACTERISTICS:						
; LENGTH: 8195 base pairs						
; TYPE: nucleic acid						
; STRANDEDNESS: double						
; TOPOLOGY: linear						
US-08-961-527-94						
Query Match						
Best Local Similarity 100.0%; Score 2388; DB 4; Length 8195;						
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;						
Qy	1	TTCTTACGAGTTGGAGCTGTATCAAGCTTAGACCGTTAAGGAAATATCGTGTTCCTA	60	1081	CCCGCAACCTGCACCAATCTTTAAATAAGACTCAAAATCTTCTTTTGGTTAGTCAGCTGGT	1140
Db	3053	TTCTTACGAGTTGGAGCTGTATCAAGCTTAGACCGTTAAGGAAATATCGTGTTCCTA	3112	4133	CCCGCAACCTGCACCAATCTTTAAATAAGACTCAAAATCTTCTTTTGGTTAGTCAGCTGGT	4192
Qy	61	TATGATGAAACAAAGCAGCCCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120	1141	ACGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCGATCTCTGTTATGTTTTC	1200
Db	3113	TATGATGAAACAAAGCAGCCCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	3172	4193	ACGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCGATCTCTGTTATGTTTTC	4252
Qy	121	CGGTGAAGGAATCAATGCTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	180	1201	GAAAGATTTACCATCTGAAACTGTTTAAATAATCTTCAAGAGCAAGTTATCAAAACAAGAGAG	1260

Db 4253 GAAAGATTTACCATCTGAAACCTGTTAAAACTCTTGAAGCAAGTTATCAAAAACAAGAG 4312
QY 1261 TGTTCACACACTTTAACTGCTAAAGAGAAATGTTGCTCCTCGTGACCAAGATTTTA 1320
Db 4313 TGTTCACACACTTTAACTGCTAAAGAGAAATGTTGCTCCTCGTGACCAAGATTTTA 4372
QY 1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGNAATAAGGGTCG 1380
Db 4373 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGNAATAAGGGTCG 4432
QY 1381 TAATTCGATTTCCAGCCCTTAGACAAATTTAGAACGCTTGAATGATGAATGACTAA 1440
Db 4433 TAATTCGATTTCCAGCCCTTAGACAAATTTAGAACGCTTGAATGATGAATGACTAA 4492
QY 1441 TAAAGAAAAATTCGTAGATGATTTATTTGGCATTCCTAGACCAATTAACCCATCCAGAGC 1500
Db 4493 TAAAGAAAAATTCGTAGATGATTTATTTGGCATTCCTAGACCAATTAACCCATCCAGAGC 4552
QY 1501 ACTTGGCAAAACCAATTTCTCAAAATTTAGTATATCTGAAGACGAAGTTGCTATTGCTCAATT 1560
Db 4553 ACTTGGCAAAACCAATTTCTCAAAATTTAGTATATCTGAAGACGAAGTTGCTATTGCTCAATT 4612
QY 1561 AGCTGATAAGTATACACGCTCAGATGTTACATTTTGTGTAACATGATATATCAGTGA 1620
Db 4613 AGCTGATAAGTATACACGCTCAGATGTTACATTTTGTGTAACATGATATATCAGTGA 4672
QY 1621 TGAAGGAGATGCATATGTTAACGCTCATATGSGCCATAGTCACTGGATTGGAAAGATAG 1680
Db 4673 TGAAGGAGATGCATATGTTAACGCTCATATGSGCCATAGTCACTGGATTGGAAAGATAG 4732
QY 1681 CTTTCTGTGTAAGGAAAAAGTTGCGCTCAAGCTATACCTAAAGAAAAAGTATCCTACC 1740
Db 4733 CTTTCTGTGTAAGGAAAAAGTTGCGCTCAAGCTATACCTAAAGAAAAAGTATCCTACC 4792
QY 1741 TCCATCTCAGACGCGAGATGTTAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db 4793 TCCATCTCAGACGCGAGATGTTAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 4852
QY 1801 CAATCTGTGTAAGGAAAAAGTATCCTGCTGTTTCCATTAAGGATCAATTACCATATAT 1920
Db 4853 CAATCTGTGTAAGGAAAAAGTATCCTGCTGTTTCCATTAAGGATCAATTACCATATAT 4972
QY 1921 TAAATTTGCTTGTGATGATCACAATACAAAGCTCAAAATGGCTATACCTTGGAGA 1980
Db 4973 TAAATTTGCTTGTGATGATCACAATACAAAGCTCAAAATGGCTATACCTTGGAGA 5032
QY 1981 TTTGTTTGGCAGATTTAAGTACTACGTAGAACACCTGACGACGTCACATTTCTAATGA 2040
Db 5033 TTTGTTTGGCAGATTTAAGTACTACGTAGAACACCTGACGACGTCACATTTCTAATGA 5092
QY 2041 TGGATGGGCAATGCGAGTGAGCATGTTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100
Db 5093 TGGATGGGCAATGCGAGTGAGCATGTTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 5152
QY 2101 TAAGAACTTCAAGCGGATGAAGCCAGTGAAGGAAACACCTGAGCCAGAGTCCC 2160
Db 5153 TAAGAACTTCAAGCGGATGAAGCCAGTGAAGGAAACACCTGAGCCAGAGTCCC 5212
QY 2161 TCAAGTAGAGATGTAAGAAAGTGAAGCCCACTCAAGAGAGAGATTTTCTGCGAA 2220
Db 5213 TCAAGTAGAGATGTAAGAAAGTGAAGCCCACTCAAGAGAGAGATTTTCTGCGAA 5272
QY 2221 AGTAAGGATTTAGTCTGTAAGAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATAA 2280
Db 5273 AGTAAGGATTTAGTCTGTAAGAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATAA 5332
QY 2281 TTTGACTCTTCAAAATTTGGATTAACATAGTATCATGGCAGAGCAAGAAATTAATTGC 2340
Db 5333 TTTGACTCTTCAAAATTTGGATTAACATAGTATCATGGCAGAGCAAGAAATTAATTGC 5392

QY 2341 GTTGTAAAAGGAAGTAATCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 5393 GTTGTAAAAGGAAGTAATCTTCATCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 5

US-09-468-656A-11
; Sequence 11, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-11

Query Match 41.5%; Score 991; DB 4; Length 2531;
Best Local Similarity 65.7%; Pred. No. 1.5e-264;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA--TAATCGTGTTC 57
Db 60 TTCTTATGAGTTGGAGCTTACCAAGCTGGTCAAGATTAAGAGAGTCTAATCGAGTTGC 119
QY 58 CTATATAGATGGAACAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGTTAG 117
Db 120 TTATATAGATGTTGATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACCAGATGAACTCAG 179
QY 118 CAAGCTGAAGAAATCAATGCTGAGCAATCTGATCAATCAAGATTAACAGCAAGGCTATGT 177
Db 180 TAAGAGGAGGAGATCAACGCCGAAACAAATGTTTATCAAGATTAACGATCAAGGTTATGT 239
QY 178 CACTTCACATGCGCACCACTATCATTTATTAACAATGTAAGGTTCCCTTATGACGCTATCAT 237
Db 240 GACCTCTCATGAGACCAATTATCATTTACTATTAATGCAAGGTTCCCTTATGATGCCATCAT 299
QY 238 CAGTGAAGAAATTAATCATGAGGATCCAAACTATAGCTTAAAGATGAGGATATTGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAAATTTATCAGTTGAAGGATTCAGACATTGTCAA 359
QY 298 TGAGGTCAAGGCTGGATATGTTATCAAGGTAGATCGAAATACTATGTTTACCTTAAGCA 357
Db 360 TGAATCAAGGTTGGTTATGTCATTAAGGTAACCGGTAATTAATTAATGTTTACCTTAAGCA 419
QY 358 TGCTGCCCAACCGGATAACGTCCTGACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db 420 TGCRCGTATCGGATAATATTTCCGACAAAAGAGAGATTAACCGTCAGAGCAGCAAGC 479
QY 418 TAGTCAACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTGCTGTGCTGGCAGCTTC 477
Db 480 CAGTCAATATCAT-----AACTCAAGAGCAGATAATGCTGTTGCTGACGACGAGC 530
QY 478 GCAAGGAGCTTACTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGA 537
Db 531 CCAAGGAGCTTATCAACGAGATGAGGATATCTTCAATGCATCTGATATCATTTAGGA 590
QY 538 TACTGCTGATGCTTATATCGTTTCTTCATGAGATCAATACCAATTAACATTCTTAAGAAATGA 597
Db 591 CACGGGTGATGCTTATATCGTTTCTTCACGGCGACCAATTAACATTACATTCTTAAGAAATGA 650

QY 598 GTTATCAGCTAGGAGTTGGCTGCTCAGAGCGCTTCTCTATCTGGTCGAGAAATCTGTC 657
 Db 651 GTTATCAGCTAGGAGTTAGCTGCTGAGAGCGCTTATGGAAATGG----- 695
 QY 658 AAATTCAAGAACCTTATCGCGACAAATAGCGATAACACTTCAAGAACAAACTGGGPACC 717
 Db 696 -GAAGCAGGATCTCGTCTCTCTCAAGTTCTAGTTATATGAAATCCAGCTCAACCAA 754
 QY 718 TTCTGTAAAGCAATCCAGGAATACTAAATATCTAAACAGCAACACAGCACTAACAG 777
 Db 755 GATTGTGAGAGAACCAATCT-----GACTGTCTCACTCAACTTA 794
 QY 778 TCAAGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTTCAAACTGCTCTTT 837
 Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTTAACCTTT 854
 QY 838 GAGTCAACAGCATGTGAATCTGATGGCTTTGCTTTGATCCAGCACAAATCAAGTGG 897
 Db 855 ATCAGACGCCATGTGGAATCTGATGGCTTTATTTTCGACCGCAAAATCAAGTGG 914
 QY 898 AACAGCTAGAGTGTGGAGTGCACACAGGAGATCATACCACCTCATCCCTTACTCTCA 957
 Db 915 AACGCCAGAGGTGTAGTGTCTCTCATGTGTAAACATTAACCATTTATCCCTTTATGAACA 974
 QY 958 AATCTGTAATTGGAAGAACGAATCGCTGCTGATTTATTTCCCTTCTGTTATCGTTCAAAACCA 1017
 Db 975 AATGCTGAATTGGAAACGAATGCTGCTGATTTATTTCCCTTCTGTTATCGTTCAAAACCA 1034
 QY 1018 TTGGGTACAGATTTCAAGGCCAGAACCAACAGTCCACACCGACTCCGGAACTTAGTCC 1077
 Db 1035 TTGGGTACAGATTTCAAGGCCAGAACCAACAGTCCACACCGACTCCGGAACTTAGTCC 1094
 QY 1078 AGGCCGCAACTGCAACCAATCTTAAATAGACTCAATTTCTTTTGGTTAGTCACT 1137
 Db 1095 AAGTCCGCAAGCTTCAAGCAATCCAAATGATGGGAA-----ATTGGTCAAGAGC 1148
 QY 1138 GGTAGCAAGTGGGAGAGATATGTATTCGAGAAAGGCGATCTCTCGTTATGCTTT 1197
 Db 1149 TGTTGCAAGTGGGAGATATGTATTTGAGGAGATGAGTTTCTCGTTATATCCC 1208
 QY 1198 TCGCAAGATTTACCATCTGAATCTTAAATCTTGAAGCAAGTTATCAAAACAGA 1257
 Db 1209 AGCCAAGATCTTTCAAGCAAGACAGCAGGAGGATTTGATGCAAACTGGCCAGCAGA 1268
 QY 1258 GAGTGTTCACACATTTAACTGCTTAAAGAGAAATGTTCTCTCGTCCAGCAGAAAT 1317
 Db 1269 AAGTTTATCTATAAGTGGAACTTAAGAAACCTGACCTCCATCTAGTGAATCGAGAAT 1328
 QY 1318 TTATGATAAGCATATAATCTGTTAACTGAGGCTCTATAAGGCTTTGTTGNAATAAGGG 1377
 Db 1329 TTACATTAAGCTTATGACTTACTAGCAAGATTTCAACAGATTTACTTGTATAATAAGG 1388
 QY 1378 TCGTAAATCTGATTTTCAAGCCTTATAGCAAAATTTAGAACGCTTGAATGAAATCGAC 1437
 Db 1389 TCGCAAGTTGATTTTGAAGCTTTGATTAACCTGTTGGAACGACTCAAGGATGTCCTAAG 1448
 QY 1438 TAATAAGAAATTTGGTAGATGATTTATGTCNTTCTAGCACCAATTTACCCATCCAGA 1497
 Db 1449 TGATAAGTCAAGTAGTGGAGATATTTCTGCTTCTTAGTCCGATTCGTCATCCAGA 1508
 QY 1498 GCGACTTGGCAACCAAAATCTTCAATTTGAGTATATCTGAAGCAGAGTTTGGTATGCTCA 1557
 Db 1509 ACGTTTAGGAAACCAATGCGCAATTTACCTACACTGATGATGATGATGATGATGATGAT 1568
 QY 1558 ATTAGCTGATAGTATACAGCTCAGATGTTTACATTTTGTGATGACATGATATATCAG 1617
 Db 1569 GTTGCAGGCAAGTACACAGCAGAGACGCTTATATCTTTGATCTCTCGTGAATACCCAG 1628
 QY 1618 TCATGAGGAGATGATATGTAACGCTCTATGAGGCTTATGAGGCTTACTGATGATGATGATGAT 1677
 Db 1629 TGATGAGGGGATGCTTATGTAACCTTCAATATGATGATGATGATGATGATGATGATGATGAT 1688
 QY 1678 TAGCCCTTCTGATAGGAAAGTTGAGCTCAAGCTTATATCTAAAGAAAGGATATCTCT 1737

Db 1689 TAGTTGTCTGAAGCTGAGAGAGCGGAGCCCGCTTATGCTTAAAGAGAAAGTTTGAC 1748
 QY 1738 ACCTTCATCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTAT 1797
 Db 1749 CCCTCTCTGACAGACCATCAGGATTCAGGAATATCTGAGCAAAAGGAGCAGAGCTAT 1808
 QY 1798 TTACAATCTGTGTGAAGGGGAAACGAAATTCACCTCGTTTCGACTTCCATATATGGTTGA 1857
 Db 1809 CTACACCGMGTGAAGCAGCTAAGAGGTGGCCACTTGTATGCTTATGCTTACATCTTCA 1868
 QY 1858 GCATACAGTTGAGGTTAAACCGGTATTTGATTTATTCCTCATAGAGATCATTAACCATAA 1917
 Db 1869 ATATCTGTAGAAGTCAAAACCGTGTATTAATCATACCTCATTAATGATTAACCATTA 1928
 QY 1918 TATTAAATTTGCTTGGTTTGTATGATCATACATACAAAGCTCCAAATGGCTATATACCTTGA 1977
 Db 1929 CATCAAAATTTGAGTGGTTTGAAGAGCCCTTATGAGCAGCTTAAGGGGTATATCTTGA 1988
 QY 1978 AGATTTCTTTGCGACGATTAAGTACTAGTAGAACACCCCTGAGCAAGCGTCCACATTTCTAA 2037
 Db 1989 GGATCTTTTGGCGACTCTCAAGTACTATGTGAAACATCCAAACCGAAGCTCGGCAATTCAGA 2048
 QY 2038 TGATGATGGGCGCATCCAGTGAAGTGTGTTAGGCAAGAAAGACCCACAGTGAAGATCC 2097
 Db 2049 TAATGTTTGGTAAACGCTAGCGACCATGTTCAAAAGAAACAAAAATGGTCAAGCTGATAC 2108
 QY 2098 AAATAAGCACTTCAAAAGCGGATGA----- 2121
 Db 2109 CAATCAAAACGAAACCAAGCAGGAGGAAACCTTCAGACAGAAACCTGAGGAGAAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCCTCGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAACCAACAGAGGAAAC 2228
 QY 2131 AGAGAAACACTCTCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAGTAGAGCCCA 2190
 Db 2229 AGAAGATTCAGCAGGAGAAATCAGAAAGAACCTTCAGTGCAGACTGAAAGGTTGAAGAAA 2288
 QY 2191 ACTCAAGAACGACAGAGTTTCTTGGCAAGTAAACGATTTCTAGTCTGAAAGGCAATGC 2250
 Db 2289 ACTGAGAGGCTGAGATTTACTTTGGAAATCCAGGATCCAATTAATCAAGTCCAATGC 2348
 QY 2251 AACGAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAATATTGATTAACATAG 2310
 Db 2349 CAAAGAGACTCTCAGGATTAATAATAATTTACTATTGCGACCCAGGACCAACATAC 2408
 QY 2311 TATCATGCGACAGCAGAAATTTACTTGGCTTGTAAAGGAGATTA 2358
 Db 2409 TATTATGCGAGAGCTGAAACAACTATTGGCTTTATTAAAGGAGAGTAA 2456

RESULT 6
 US-09-468-656A-5
 ; Sequence 5, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Adamou, John E.
 ; APPLICANT: Johnson, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: 1999-12-02
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae

US-09-468-656A-5

Query Match 41.4%; Score 990; DB 4; Length 2531;
 Best Local Similarity 68.1%; Pred. No. 2.8e-264;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 1 TTCTTACGAGTGGGACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTGTTC 57
 DB 60 TTCTATGAACTTGGTGGTCAACCAAGCTGGTCAAGGTTAAGAAAGAGTCTAATCGAGTTTC 119

QY 58 CTATATAGATGGAAACACAGCGACGCAAAAACCGAGAAATTTGACTCCTGATGAGGTTAG 117
 DB 120 TTATATAGATGGTATCAGGCTGGTCAAAAGGCGAGAAACTTGACACCAAGTGAAGTCAG 179

QY 118 CAAGCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGT 177
 DB 180 TAAAGGGAGGGGATCAACCGCGAACAATTCGTCATCAAGATTACGGATCAAGGTTATGT 239

QY 178 CACTTCACATGGCGACCACTATCATTTATTAACAATGGTAAGGTTCCCTTATGACGCTATCAT 237
 DB 240 GACCTCTCATGGAGACCAATTATCATTTACTATATATGCAAGGTCCTTATGATGCCATCAT 299

QY 238 CAGTGAAGATTTACTCATGAAAGATCAAACTATAGCTTAAAGATGAGAGATTTGTTAA 297
 DB 300 CAGTGAAGAGCTCCTCATGAAAGATCGAATTTATCAGTTGAAGGATTCAGACATTTGTCAA 359

QY 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 357
 DB 360 TGAATCAAGGTGGTATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 419

QY 358 TGCTGCCCGCGGATAACGTCGCTCAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
 DB 420 TGCAGCTCATCGGATAATATTCGGACAAAAGAGAGATTAAACGTCAGAGCAGGAACA 479

QY 418 TAGTCAACATCGTGAAGGTGGAATCAAGAAACGATGGTGGTGGCTTGGCATTGCAAGTTC 477
 DB 480 CAGTCAATATCAAGGGGGTGGTCT-----AACGATCAAGCAGTAGTTGCAAGCCAGAGC 533

QY 478 GCAAGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCATGATATCATAGGA 537
 DB 534 CCAAGGACGCTATACACGATGATGTTATATCTTCATGATCACTGATATCATGAGGA 593

QY 538 TACTGTGATGCTTATATCGTTCCTCATGGAGATCAATACCAATTAATTCCTAAGAAATGA 597
 DB 594 CACGGGTGATGCTTATATCGTTCCTCACGCGACCAATTACCAATTACATTCCTTAAGATGA 653

QY 598 GTATACAGTACGAGTGGCTGCTGCAGAACGCTTCCTATCTGGTTCGAGGAATCTGTC 657
 DB 654 GTATACAGTACGAGTGGCTGCTGCAGAACGCT----- 688

QY 658 AAATTCAGAACTATCGCCGACAAAATAGCGATACACTTCAAGAACAACTGGGTACC 717
 DB 689 -----ATTGGAATGGGAAGCAGGGATCTGCTCTCTTCAAGTTCTAGTTTAAATGC 740

QY 718 TTCTGTGAAGCAATCCAGGAACCTACAAATACTAACACAGCAACAACAGCAACACTAACAG 777
 DB 741 AAATCCAGCTCAACCAAGATTGTCAAGAAACCAACATCTGATGTCCTCACTCAACTATCA 800

QY 778 TCAAGCAAGTCAAGTAAATGACATTTGATGATGCTCTTTGAAACAGCTCTACAAATCGCTTT 837
 DB 801 TCA---AAATCAAGGGGAAAAAATTTCAAGGCTTTTACGTTGAATTTGATGCTAAACCCCTT 857

QY 838 GAGTCAACGACATGTAGAATCTGATGGCTGTGTTGATCCAGACAAATCACAGTCG 897
 DB 858 ATCAGAACCCATGTGGAATCTGATGGCCCTTATTTTCGACCCAGCGCAAAATCACAGTCG 917

QY 898 AACAGCTAGAGGTGTGAGTGCCACACGAGATCAATTACCACTTCATCCCTTACTCTCA 957
 DB 918 AACCGCCAGAGGTAGTGTCCCTCATGTGAACCAATACCACTTTATCCCTTATGAACA 977

QY 958 AATGCTGAAATTTGGAAGAACGAATCGCTGATATTTATCCCTTGGTATCGTTTCAAAACA 1017
 DB 978 AATGCTGAAATTTGGAAGAACGAATTTGCTGATTTATTCCTCTTGGTTATCGTTTCAAAACA 1037

QY 1018 TTGGTACCAGATTCAAGGCCAGAACCAAGTCCACCAAGCTCCGAACTCGGAACTTAGTCC 1077
 DB 1038 TTGGTACCAGATTCAAGGCCAGAACCAAGTCCACCAAGCTCCGAACTCGGAACTTAGTCC 1097

QY 1078 AGCCCGCAACCTCGACCAATCTTAAATFAGACTCAAAATTTCTC-----TTT 1125
 DB 1098 AAGTCGCAACCTCGACCAATCTCAACAGCTCCAAAGATCCAAATTCGATGAGAAAT 1157

QY 1126 GGTATGAGTGGTACGAAAGTTTGGGGAAGGATATGTATTCGAAAGAAAGGCGATCTC 1185
 DB 1158 GGTCAAGAAGCTGTTTCGAAAGTAGGCGATGTTATGCTTTGAGGAGAAATCGAGTTTC 1217

QY 1186 TCGTTATGCTTTGCGAAAGATTTACATCTGAAACTGTTTAAAAATCTTGAAGAGCAAGTT 1245
 DB 1218 TCGTTATATCCAGCCAGGATCTTTTTCAGCAGAAACAGCAGCAGGATTTGATAGCAAACT 1277

QY 1246 ATCAAAACAAGAGAGTGTTCACACACTTTTAACTGCTAAAAAGAAATGTTGCTCCTCG 1305
 DB 1278 GGCAGAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAACTGACCTCCATCTAG 1337

QY 1306 TGACCAAGAATTTTATGATAAAGCATATAATCTGTTAACTAGGCTCATAAAGCCTTGT 1365
 DB 1338 TGATCGAGAATTTTCAATAAGGCTTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397

QY 1366 TGAATAAAGGTCGTAATCTGATTTTCCAAAGCTTAGCAAAATTTAGAACGCTTGAA 1425
 DB 1398 TGATAAAGGTCGCAAGTTGATTTTGGGCTTTGGAATTAACCTGTTGGAACGACTCAA 1457

QY 1426 TGATGAATCGACTAATAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTGACCAAT 1485
 DB 1458 GSNATGCCAAGTGATTAAGTCAAGTTAGTGGATGATATCTTGCTCTTTAGCTCCGAT 1517

QY 1486 TACCCATCCAGAGCGACTTGGCAAAACCAAAATCTCAAAATTTAGTATATCTGAAGACGAA 1545
 DB 1518 TCGTATCCAGAACGTTTAGGAAAAACCAATGCGCAATTAACCTACACTGATGATGAT 1577

QY 1546 TCGTATTTGCTCAATAGCTGATAAGTATACAAAGTCAAGCTCAGATGTTTATTTGATGAACA 1605
 DB 1578 TCAAGTAGCCTAAGTTGGCAGGCAAGTACACACAGAAAGCGGTTATATCTTTGATCCTCG 1637

QY 1606 TGATATAATCACTGATGAAGGAGATGCATATGTAACGCTCATATGGGCATAGTCACTG 1665
 DB 1638 TGATATAACAGTATGAGGGGATGCTTATGTAATCTCACATATGACCATAGCACCTG 1697

QY 1666 GATTTGAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATATACTAAAGA 1725
 DB 1698 GATTAATAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGA 1757

QY 1726 AAAAGGTATCCTACCTCCATCTCCAGACGCGATGTTAAAGCAAAATCAACTGGAGATAG 1785
 DB 1758 GAAAGTTTGAACCCCTCTCTTCAGACAGCCATCAGGATTTCAAGAAATFACCTGAGGCAAAAG 1817

QY 1786 TGCAGCAGCTATTTACAAATCGTGTGAAAGGGGAAAAACGAAATTCACCTCGTTCGACTCC 1845
 DB 1818 ASCAGAGCTATCTACACCGCTGGAAGCAGCTAAGAGGTGCCACTTGATCGTATGCC 1877

QY 1846 ATATATGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTTATTTCTCATAAGGA 1905
 DB 1878 TTACAATCTTCAATATATCTGTAGAAGTCAAAAAACGGTAGTTTAAATCATACCTCATATGA 1937

QY 1906 TCATTTACCATATATTAATTTGCTTGGTGGTGGATGATCACATACAAAGCTCCAAATGG 1965
 DB 1938 CCAATACCATTAACATCAAAATTTGAGTGGTTTGAACGAGCCCTTTATGAGGCCACTTAAGGG 1997

QY 1966 CTATACCTTGAAGATTTGTTTGGCAGCATTAAGTACTAGCTAGAACACCTCGAGCAAGC 2025
 DB 1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTTCGAACATCCAAAGCAAGC 2057

QY 2026 TCCACATTTAATGATGGATGGGCAATGCCAGTGGAGCATGTTTGGCAAGAAAGACCA 2085
 DB 2058 TCCGCAATTCAGATAATGGTTTGGTAAACGCTAGCGACCATGTTTCGTAATAAATAAGGTAGA 2117


```
QY 1366 TGAATAAAGGTCGTAATCTGATTTCCAAAGCCTTAGACAAATTAATTAGAACGCTTCAA 1425
Db 1339 TGATAATAAGGTCGCAAGTTGATTTGAGCCTTTGGATTAACCTGTTGGAAACGACTCAA 1398
QY 1426 TGATGAATCGCACTAATAAGAAAAATTTGGTAGATGATTTAATGCGATTCCTAGACCAAT 1485
Db 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATTCTTCCCTTTTGTAGCTCCGAT 1458
QY 1486 TACCATTCCAGAGGACTTCGCAACCAAAATTTCTCAAATTCAGTATGATGAGAGCAAGT 1545
Db 1459 TCGTATCCAGAGGTTTAGGAAACCAAAATGCGCAAAATTAACCTACACTGATGATGAGAT 1518
QY 1546 TCGTATTCCTCAATTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605
Db 1519 TCAAGTAGCCAAAGTTGCGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1578
QY 1606 TGATAATAAGGTCGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
Db 1579 TGATAATAAGGTCGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1638
QY 1666 GATTGGAAGAGATAGCTTTCTGATAAGGAAAGTTGCGAGCTCAAGCTCAAGCTTACTAAAGA 1725
Db 1639 GATTAAAGAGATAGTTTGTGAAGCTGAGAGCGCGAGCGCCAGGCTTTATGCTAAAGA 1698
QY 1726 AAAAGGTATCTTACCTCCATCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTCGAGATAG 1785
Db 1699 GAAAGGTTTGACCCCTCTTCGACAGACCATCAGATTTCAGGAAATTAAGGCAAAAGG 1758
QY 1786 TGCAGAGCTATTTACAACTGCTGAAAGGGGAAACGAATTCCTGCTGCTGACTTCC 1845
Db 1759 AGCAGAGGCTATCTCAAAACCGCTGAAAGCAGCTAAGAGGTCCTGATGATGCTGATGCC 1818
QY 1846 ATATATGTTTGACATACAGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1905
Db 1819 TTCAATCTTCAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1878
QY 1906 TCATTACCAATATTAATTTCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1965
Db 1879 CCATTACCAATACATCAATTTGAGTGGTTGAGAGGCTTTNTGAGGCACTAAGGG 1938
QY 1966 CTATACCTTGGAGATTTGTTTGGAGATTAAGTACTACGTAGAACACCTGACGAACG 2025
Db 1939 GTATACCTTGGAGATTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 1998
QY 2026 TCACATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
Db 1999 TCGCATTCAGATAATGTTTGGTACGCTAGCGACCATGTTTCAAGAAACCAAAATGG 2058
QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
Db 2059 TCAAGCTGATACCAATCAAAACGGAAGAAACCAAGCGAGAGAGAAACCTCAGACAGAAACC 2118
QY 2146 TGAGCCAGAGTCCCTCAAGTAGAGA 2171
Db 2119 TGAGGAAGAAACCCCTCGAGAAGAGA 2144
```

RESULT 8

```
US-09-536-784-65
; Sequence 65, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-09-536-784-65
```

```
Query Match 41.3%; Score 987.6; DB 4; Length 2290;
Best Local Similarity 67.7%; Pred. No. 1.2e-263;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;

QY 1 TTCTTACGAGTTCGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTC 57
Db 4 TTCTTATGAACCTTGGTCGTCAACCAAGCTGTCAAGTTAAGAAAGAGTCTAATCGAGTTTC 63
QY 58 CTATATAGATGGAACAAACAGCAGCGCAAAACCGAGAAATTCACCTCCTGATGAGGTTAG 117
Db 64 TTATATAGATGTTGATCAAGCTGGTCAAAAGGCGAGAAACCTTGACACCAAGTGAAGTCAG 123
QY 118 CAAGCTGTAAGGAATCAATGCTGAGCAAAATCGTCAACAGATAACAGACCAAGCCTATGT 177
Db 124 TAAGAGGGAGGGATCAACGCCGACAAATNGTNAICAAAGATTACGATCAAGGTTATGT 183
QY 178 CACTTCATATGCGACCACTATCAATTAATCAATGTTAAGGTTTCCTTATGACGCTATCAT 237
Db 184 GACCTCTCATGAGACCAATTAATCAATTAATTAATGCAAGGTTTCCTTATGATGCCATCAT 243
QY 238 CAGTGAAGAAATTAATCAATCAAGAGATCCAAACTATTAAGCTTAAAGATGAGGATATTGTTAA 297
Db 244 CAGTGAAGAGCTCCTCATGAAAGATCCGAAATTAATCAGTTGAAGGATTCAGACATTTGTCAA 303
QY 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGA 357
Db 304 TGAATCAAGGGTGGTATGTTCAATTAAGGTAACCGTAATTAATCTATGTTTACCTTAAGGA 363
QY 358 TGCTGCCACCGCGATAACGTCCTGATCAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db 364 TGCAGCTCATGCGGATAATATTCGGACAAAAGAGAGATTAACGTCGACAGCAGAAACG 423
QY 418 TAGTCAACATCGTGAAGGTGGAACCTCCAAAGAACCATGTTGCTGTTGCTGCGACGTTTC 477
Db 424 CAGTCATATCAT-----AACTCAAGAGCAGATAATGCTGTTGTCGACCCAGAGC 474
QY 478 GCAAGGACGCTATCTACAGATGATGTTTATCTTTAATGTTCTCTGATATCATAGAGA 537
Db 475 CCAAGGACGTTATACAAAGGATGATGTTATCTTCAATGATCTGATATCATTTAGAGA 534
QY 538 TACTCGTATGCTTATATGTTTCCTCATGAGATCAATACCATTAATCCCTTAAGAAATGA 597
Db 535 CACGGGTGATGCTTATATGTTTCCTCAACGGGACCATTAACCATTAATCCCTTAAGAAATGA 594
QY 598 GTTATCAGTAGGAGTTGGCTGTCGAGAAAGCCTTCTTATCTGTCGAGGAAATCTGTC 657
```

Db	595	GTATACGCTAGCGAGTATGCTGCTGAGAGCCT-----	629
Qy	658	AAATTCAGAACCTTATCCCGACAAAATAGCGATAACATTCAGAAACAACTGGGTACC	717
Db	630	-----ATTGGAATCGGAAGCAGGAGATCGTCCCTTCTCAAGTTCTAGTTATAATGC	681
Qy	718	TTCTGTAAGCAATCCCGAACTACAAATACACACAGCAACAAACACACACACACACAG	777
Db	682	AAATCCAGCTCAACACAGATTTGTGAGAGAACCAAAATCTGACTGTCTCACTCAACTTATCA	741
Qy	778	TCAGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAACACAGCTCTACAAACTGCTTTT	837
Db	742	TCA-----AAATCAAGGGGAACATTTCAAGCTTTTAGTGGAATTTGATGCTAAACCTT	798
Qy	838	GAGTCAACGACATAGAACTCTGATGCTTGTCTTTGATGACGACAAATACACAAAGTCG	897
Db	799	ATCAGAACGCGCATGTGAAATCTGATGCTTTTATTTTCGACCGGCAAAATACAAAGTCG	858
Qy	898	AACAGCTAGAGCTGTCAGTGCACACGAGAGATCATACCACTTCATCCCTTACTCTCA	957
Db	859	AACCGCAGAGGTAGTGTCCCTCATGTGTAAACATTAACCATTTATCCCTTATGACAA	918
Qy	958	AATGCTGAAATTGGAAGAACGNAATCGTCTGATTAATTTCCCTTCTGTTATCTTCAAAACA	1017
Db	919	AATGCTGAAATTGGAAGAACGNAATGCTGATTAATTTCCCTTCTGTTATCTTCAAAACA	978
Qy	1018	TTGGGTACAGATTAAGGCGCAGAACAAACCAAGTCCCAACCGACTCCGGAACCTAGTCC	1077
Db	979	TTGGGTACAGATTAAGGCGCAGAACAAACCAAGTCCCAACCGACTCCGGAACCTAGTCC	1038
Qy	1078	AGGCGCGCACTGCACCAAAATCTTAAATAGACTCAAAATCTTTC-----TTT	1125
Db	1039	AAGTCCGCAACTGCACCAAAATCTTAAATAGACTCAAAATCTTTC-----TTT	1098
Qy	1126	GGTTAGTCAAGCTGTTGGAAGGAGGATGATGATTTGGAAGAAAGGCGATCTC	1185
Db	1099	GGTCAAGAGAGCTGTTGGAAGGAGGAGGATGATGATTTGGAAGAAAGGCGATCTC	1158
Qy	1186	TGTTATGCTTTTGGGAAGATTTACCAATCTGAACTGTTAAATCTTGAAGCAAGTT	1245
Db	1159	TCGTTATATCCAGCAAGGATCTTTTCAGCAGAAACACAGCAGGAGCATTTGATCAAACT	1218
Qy	1246	ATCAAAACAGAGAGTGTTCACACATTTTAACTGTAAGAAAGAAATGTTGCTCCCTCG	1305
Db	1219	GGCCAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACAGTCCCTCCATCTAG	1278
Qy	1306	TGACCAAGAAATTTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGT	1365
Db	1279	TGATCGAGAAATTTTACATAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACT	1338
Qy	1366	TGNAATTAAGGCTGTAATTTCTGATTTCCAGCCTTAGACAAATTTTAGACGCTTGAA	1425
Db	1339	TGATAATAAGGCTGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAAACGACTCAA	1398
Qy	1426	TGATCAATGCTACTATAAGAAATTTGATGATGATTTTGGGCTTTCTAGCAACCAAT	1485
Db	1399	GGATGTCNCAAGTATAAGTCAAGTTAGTGGAGATTTCTGCTTCTTAGCTCCGAT	1458
Qy	1486	TACCCATCCAGAGCAGTGGCAACCAAAATTTCTCAAAATTTAGTATATCTGAAAGCAAGT	1545
Db	1459	TCGTATCCAGAACGTTTAGGAAACCAAAATGCGAAATTTACCTACACTGATGATGAT	1518
Qy	1546	TCGTATGCTCAATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	1605
Db	1519	TCAGTAGCAAGTTGGGAGGCAAGTACACACAGAAACGCTTATATCTTTGATCTCG	1578
Qy	1606	TCATATATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1665
Db	1579	TGATATATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1638
Qy	1666	GATTCGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1725

RESULT 9
US-08-961-527-243
; Sequence 243, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-243

Query Match 27.4%; Score 653.8; DB 4; Length 2359;
Best Local Similarity 68.3%; Pred. No. 4.3e-171;
Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

```
QY 1 TTCTTACGAGTTCGGAGCTGTATCAAGCTAGAACGGTTAAGGAAA-----TAATCGTGTTC 57
Db 939 TTCTTACGAGTTCGGAGCTGTATCAAGCTAGAACGGTTAAGGAAA-----TAATCGTGTTC 998
QY 58 CTATATAGATGCGAAGAACGAGGAGCGAAGAAACGAGAAATTCGACTCCTCATGAGGTTAG 117
Db 999 TTATATAGATGCGAGCTGTATCAAGCTAGAACGGTTAAGGAAA-----TAATCGTGTTC 1058
QY 118 CAAAGCTGAGGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGT 177
Db 1059 TAAGAGGAGGAGGATCAACCGCCGAAACAAATGTTATCAAGATTACGATCAAGGTTATGT 1118
QY 178 CACTTCACATGCGGACCACTATCATTTATTAATGTTAAGGTTCCCTATGACGCTATCAT 237
Db 1119 GACCTCTCATGAGAGCAATTCATTAATGAGCAAGGTTCCCTATGATGCCATCAT 1178
QY 238 CAGTGAAGAAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
Db 1179 CAGTGAAGAGCTCTCATGAAGATCGGAATTCAGTTGAAGGATTCAGACATTTGTCAA 1238
QY 298 TGAGGTCAAGGTTGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGA 357
Db 1239 TGAATCAAGGTTGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGA 1298
QY 358 TGCTGCCACGCGGATAACGTCGTCAAGAGAGGAAATCAATCGACAAACAAAGCA 417
Db 1299 TGAGCTCATGCGGATAATTCGGAACAAAGAGAGATTAACGTCGAGAGCAGGAACG 1358
QY 418 TAGTCAACATCGTGAAGGTGGAAGCTCAAGAAACGATGGTGTGTTGCCCTTGACAGTTC 477
Db 1359 CAGTCATATCAATCT-----CAAGAGCAGATATGCTGTTGCTGACCCAGAGC 1409
QY 478 GCAAGAGCGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db 1410 CCAAGGACGCTTATACAGCGATGATGGGTATATCTTCAATGATCTGATATCATTCAGGA 1469
QY 538 TACTGTGATGCTTATATGTTCTCATGAGATCAATACCATTAATCTTAAGATGA 597
Db 1470 CACGGTGAATGCTTATGTTCTTCAAGGTCTAGTTTATTAATGCAATCCAGCTCAACCAA 1529
QY 598 GTTATCAGCTAGCGAGTGTGGCTGCTGAGAGGCTTCTCTATCTGCTCGAGGAAATCTGTC 657
Db 1530 GTTATCAGCTAGCGAGTGTAGCTGCTGAGAGGCTTATGGAATGG----- 1574
QY 658 AAATTCAGAACTATTCGCGGCAAAATAGCGATAACATTCAGAGAAACAACTGGGTACC 717
Db 1575 -GAAGCAGGAGTCTGCTCTCTTCTTCAAGTCTAGTTTATTAATGCAATCCAGCTCAACCAA 1633
QY 718 TTCTGTAAGCAATCCAGGAACTACAACTAACACAGCAACAGCAACACTAACAG 777
Db 1634 GATTTGTCAGAGAACCAATCTGACTGT-----CACTCCAACTTA 1673
QY 778 TCAAGCAAGTCAAGATTAATGACATTAATGATAGTCTCTTGAACAGCTCAAAACGCTTT 837
Db 1674 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTTGCTTAACCCCTT 1733
QY 838 GAGTCAACAGATGATAGATCTGATGGCTTGTCTTTGATCCAGCAAAATCACAAGTCG 897
Db 1734 ATCAGAACCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCACAAGTCG 1793
QY 898 AACAGCTAGAGTGTTCGATGCGCACAGGAGATCAATTACCACTTATCCCTTACTCTCA 957
Db 1794 AACCCGACAGGTTGAGTGTCTTCCCTCATGGAACCAATTAACCACTTATCCCTTATGAAC 1853
QY 958 AATGTCGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTCCCTTCGTTATCGTTCAAACCA 1017
```

RESULT 10

```
US-08-961-083-181
; Sequence 181, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
```

;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-181

Query Match 16.1%; Score 385.4; DB 3; Length 1342;
Best Local Similarity 65.3%; Pred. No. 8.1e-97;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

Qy	37	TAAGGAAATAATCGTGTTCCTATATAGATGGAAAAAAGCGAGCGGAAACCGGAGAA	96
Db	28	TAAGGACAAATAATCGTGTTCCTATATAGATGGAAAAAAGCGAGCGGAAACCGGAGAA	87
Qy	97	TTTGACTCCTGATCAGGTAGCAAGCGGTGAGGAATCAATGCTGAGCAAAATCGTCATCA	156
Db	88	CTTGACACCAAGCAGGTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATCGTCATCA	147
Qy	157	GATTAACAGCAAGGCTATGCTACATTCATGGCGACCACTATCTATTATTAATGATGTA	216
Db	148	AATTACAGATCAGGCTATGTAACCTCACCGGTGACCACTATCTATTATTAATGATGTA	207
Qy	217	GGTTCCTTATGACGCTATCATCAGTGAAGAATTAATGCTGAGCAAAATCGTCATCA	276
Db	208	AGTTCCTTATGATGCCCTCTTTAGTGAAGAATTAATGCTGAGCAAAATCGTCATCA	267
Qy	277	AAAGATGAGGATATTTGTAATGAGGTCAAGGCTGAGATGTTATCAAGGTAGATGGA	336
Db	268	TAAAGACGCTGATATTTGTCMAATGAAGTCAAGGCTGAGTATATCATCAAGGTGAG	327
Qy	337	ATACTATGTTTACCTTAAAGATGCTGCCACCGGATACCGTCCCTACAAAGAGAAAT	396
Db	328	ATATTAATGCTTACCTGAAGATGACGCTCATGCTGATAATGTTGCAACTAAGATGA	387
Qy	397	CAATCGACAAACAAAGAGATGATCAACATCGTGAAGTGAATCGAAGTGAATCGAAG	456
Db	388	CAATCGTCAAAACAAAGAGATGATCAACATCGTGAAGTGAATCGAAGTGAATCGAAG	435
Qy	457	TGCTGTTCCTTGGACCTTGCAGGACGCTTACTACAGATGATGTTATATCTTTAA	516
Db	436	TAATGTTCTGTAGCAAGGCTTCAGGACGATATACGCAAAATGATGTTATGCTTTAA	495
Qy	517	TGCTTCTGATATCATAGAGTATCTGTGATGCTTATATCGTTCCTCATGAGATCATTA	576
Db	496	TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGCTTCTCATGAGGTCATTA	555
Qy	577	CAATTACATTCCTAAGATGATTAATGATGAGGATGCTGCTGTCGAGAGCTTCTCT	636
Db	556	TCACTACATTCCTAAGAGGATTTATCTGCTAGTGAATTAGCAGCAGTAAAGCATCT	615
Qy	637	ATCTGGTCGAGAAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATAACAC	696
Db	616	GGCTGGAAAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGCTAGTGACA	672
Qy	697	TTCAAGAACAAACTGGGTACCTTCTGTGAAGCAATCCAGAACTACAAATATCAACAAG	756
Db	673	-----TAACAGCAATCTGT	687
Qy	757	CAACACAGCAACACTAAGTCAAGCAAGTCAAAAGTAAATGACATTTGATGCTCTTGA	816
Db	688	AGCAAAAGGATCAACTAGCAAGCCAGCAAAATTAATCTGAAATCTCCAGAGCTTTTGA	747
Qy	817	ACAGCTCTACAACTGCCTTTTGAGTCAACGACATGTAGAATCTGATGCGCTTGTCTTGA	876
Db	748	GGAACTCTATGTTACCTAGCGCCACAGTTTACAGTGAATCAGATGGCTGTCTTGA	807
Qy	877	TCCAGCAAAATCAAGTCGACAGCTAGAGGTGTTGAGTGGCCACACGAGGATCATTA	936
Db	808	CCCTGCTAAGATTATCAGTCGTACACAAATGGAGTTGCGATTCCGCATGGCGACCATTA	867
Qy	937	CCACTTCATCCCTTACTCTCAATGCTGAATTTGGAAGAACGAATCGCTCGTATTATCC	996
Db	868	CCACTTTTATCTCTAGCAAGCTTTCTGCTCTTAGAAGAAAGATTCGAGAAATGGTGCC	927

Qy 997 CCTTCGT 1003
Db 928 TATCAGT 934

RESULT 11

US-09-536-784-181
; Sequence 181, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-536-784-181

Query Match	16.1%;	Score 385.4;	DB 4;	Length 1342;
Best Local Similarity	65.3%;	Pred. No. 8.1e-97;		
Matches	631;	Conservative	0;	Mismatches 276; Indels 60; Gaps 2;
Qy	37	TAAGGAAATAATCGTGTTCCTATATAGATGGAAAAAAGCGAGCGGAAACCGGAGAA	96	
Db	28	TAAGGACAAATAATCGTGTTCCTATATAGATGGAAAAAAGCGAGCGGAAACCGGAGAA	87	
Qy	97	TTTGACTCCTGATCAGGTAGCAAGCGGTGAGGAATCAATGCTGAGCAAAATCGTCATCA	156	
Db	88	CTTGACACCAAGCAGGTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATCGTCATCA	147	
Qy	157	GATTAACAGCAAGGCTATGCTACATTCATGGCGACCACTATCTATTATTAATGATGTA	216	
Db	148	AATTACAGATCAGGCTATGTAACCTCACCGGTGACCACTATCTATTATTAATGATGTA	207	
Qy	217	GGTTCCTTATGACGCTATCATCAGTGAAGAATTAATGCTGAGCAAAATCGTCATCA	276	
Db	208	AGTTCCTTATGATGCCCTCTTTAGTGAAGAATTAATGCTGAGCAAAATCGTCATCA	267	
Qy	277	AAAGATGAGGATATTTGTAATGAGGTCAAGGCTGAGATGTTATCAAGGTAGATGGA	336	
Db	268	TAAAGACGCTGATATTTGTCMAATGAAGTCAAGGCTGAGTATATCATCAAGGTGAG	327	
Qy	337	ATACTATGTTTACCTTAAAGATGCTGCCACCGGATACCGTCCCTACAAAGAGAAAT	396	

Db 328 ATATTATGCTCTACCTGAAAGATGCGAGCTCATGCTGATGATGTTGCAACTTAAGATGAAT 387
 QY 397 CAATCGACAAAACAAAGAGATAGTCAATCGTGAAGGTGAACTTCCAAGAAACGATGG 456
 Db 388 CAATCGTCAAAAACAAAGACATGTCAAAGATATGAGAAAGTTA-----ACTC 435
 QY 457 TGCTGTGCTTGGCAGCTTCCAGGAGCGCTATATCTACTACAGATGATGGTTATCTTTAA 516
 Db 436 TAATGTGCTGTAGCAAGGTCTCAGGAGCGATATACGACAAATGATGGTTATGCTTTAA 495
 QY 517 TGCTTCTGATATCATAGAGGATCTGCTGATGCTTATATCGTTCTCTCATGGAGATCATTA 576
 Db 496 TCCAGCTGATATATCGAAGATACGGGTAAATGCTTATATCGTTCTCTCATGGAGTCACTA 555
 QY 577 CAATTACATCTTAAGATGATGATATACGCTAGCGAGTTGGCTGTGCGAAGCCCTTCT 636
 Db 556 TCACGTACATTTCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 615
 QY 637 ATCTGCTCGAGGAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATAACAC 696
 Db 616 GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTCAACAGCTAGTGACAA--- 672
 QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATACTTAACACAAG 756
 Db 673 -----TACACGCAATCTGT 687
 QY 757 CAACACAGCAACACTTAACAGTCAAGCAAGTCAAAAGTAAATGATGATGATGCTCTTGA 816
 Db 688 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAATCTGAAAATCTCCAGAGTCTTTTGA 747
 QY 817 ACAGCTCTACAAACTGCTTTGAGTCAAGCATGTAGAACTGATGATGCTGCTTTTGA 876
 Db 748 GGAATCTATGATTCACCTAGCGCCCAAGTTACAGTGAATCAGATGGCTGGTCTTTGA 807
 QY 877 TCCAGCAAAATCAAGTTCGAACAGCTAGAGGTGTTCGAGTGGCCACACGAGATCATTA 936
 Db 808 CCTCTGAAGATTATCAGTCGTACACCAATGAGTTGGATTCCGATGGGACCATTA 867
 QY 937 CCATTCATCCCTTACTCTCAATGCTGAAATGGAAGCAAGAAATGCTGCTGATTTATTC 996
 Db 868 CCACCTTTATCTTACAGCAAGCTTTCTGCTTTAGAGAAAAGATTGCCAGAATGGTGCC 927
 QY 997 CCTTCGT 1003
 Db 928 TATCAGT 934

RESULT 12
 US-09-468-656A-7
 ; Sequence 7, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John B.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-7

Query Match 16.1%; Score 385.4; DB 4; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 8.4e-97;

Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAGAAATAATCGTGTTCCTATATAGATGGAAAAAAGCGACGCAAAAAACGGAGAA 96
 Db 99 TAAGACAATAATCGTGTCTCTTATGTGATGGCAGCCAGTCAAGTGCAGAAAAGTGA 158
 QY 97 TTTGACTCTGATGAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAATCGTCATCAA 156
 Db 159 CTTGACACAGACCAAGGTAGCCGAAAAGGAATTCAGGCTGAGCAATTTGTAATCAA 218
 QY 157 GATAACAGACCAAGGCTATGTCACTTTCATGCGCACCACATCATTTATTTACAATGTTAA 216
 Db 219 AATTACAGATCAGGGCTATGTAACGTCAACGTCACACATCATCTACTATATTAATGGNA 278
 QY 217 GGTTCCTTATGACGCTATCATCAGTGAAGAAATTAATCTATGAAGATCCAAATATAAGCT 276
 Db 279 AGTTCTTATGATGCCCTCTTTAGTGAAGAACTCTTTGATGAAGATCCAAATATCAACT 338
 QY 277 AAAAGATGAGGATATTTGTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
 Db 339 TAAAGACGCTGATATTTGTAATGAAGTCAAGGGTGGTTATATCATCAAGGTGATGAAA 398
 QY 337 ATACTATGTTTACCTTAAAGGATGCTGCCACGCGATTAACGTCCTGACAAAAGAGAAAT 396
 Db 339 ATATTATGCTTACCTGAAGATGACGCTCATGCTGATTAATGTTTGAATTAAGATGAAT 458
 QY 397 CAATCGACAAAACAAAGAGCATAGTCAACATGCTGAAGGTGGAATCCCAAGAAACGATGG 456
 Db 459 CAATCGTCAAAAACAAAGAACATGTCAAAGATAATGAGAAGTTA-----ACTC 506
 QY 457 TGCTGTGCTTGGCAGCTTCCAGAGGACGCTATATCTACAGATGATGTTATCTTTAA 516
 Db 507 TAATGTTGCTGTAGCAAGGTCTCAGGAGCATATACGACAAATGATGGTTATGTTTAA 566
 QY 517 TGCTTCTGATATCATAGAGGATCTGCTGATGCTTATATCGTTCTCTCATGAGATCATTA 576
 Db 567 TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGTTCTCTCATGGAGTCACTA 626
 QY 577 CCATTACATCTTAAGATGATGATCAGCTAGCGAGTTGGCTGTGCTGAGAAAGCCCTTCT 636
 Db 627 TCACGTACATTTCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 686
 QY 637 ATCTGCTCGAGGAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATAACAC 696
 Db 687 GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGCTAGTGACAA--- 743
 QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATACTTAACACAAG 756
 Db 744 -----TACACGCAATCTGT 758
 QY 757 CAACACAGCAACACTTAACAGTCAAGCAAGTCAAAAGTAAATGATGATGATGCTCTTGA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAATCTCCAGAGTCTTTTGA 818
 QY 817 ACAGCTCTACAACTGCTTTGAGTCAAGCATGTCAGAAATGATGATGCTGCTTTTGA 876
 Db 819 GGAATCTTATGATTCACCTAGGCGCCAACGTTACAGTGAATCAGATGGCTGGTCTTTGA 878
 QY 877 TCCAGCAAAATCAAGTTCGAAACAGCTAGAGGTGTTGAGTGGCCACACGAGATCATTA 936
 Db 879 CCCTGTGAAGATTATCAGTCTGTAACCAATGAGGTGCGATTCCGATGGGACCATTA 938
 QY 937 CCATTCATCCCTTACTCTCAATGCTGAAATGGAAGAAAGAAATGCTGCTGATTTATTC 996
 Db 939 CCACCTTTATCTTACAGCAAGCTTTCTGCTTTAGAGAAAAGATTGCCAGAATGGTGCC 998
 QY 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 13
 US-08-961-527-192/c


```
; Sequence 192, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Hp Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-192

Query Match 16.1%; Score 385.4; DB 4; Length 6867;
Best Local Similarity 65.3%; Pred. No. 2e-96;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATAATCGTGTTCCTATATAGATGGAACAAAGCAGCGCAAAACCGAGAA 96
Db 6687 TAAGGACAAATACGCTGTCTTATGTGGATGGCAGCGCAGTCAAGTCAAGAAAGTGAAA 6828

QY 97 TTGTGACTCCTGATGAGGTTAGCAAGCGGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAA 156
Db 6627 CTTGACACCAAGCAGGTTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATGTAATCAA 6568

QY 157 GATAACAGACCAAGGCTATGTCATTCACATGGCGGACCACTATCAATATTACATGCTAA 216
Db 6567 AATTACAGATCAGGCTATGTAAGCTCACAAGGTCAGGCTATCAATATTACATGCTAA 6508

QY 217 GGTTCCTTATGACGCTATCATCACTGAGGAATTAATCAAGAAATCAAACTATAAGCT 276
Db 6507 AGTTCCTTATGAGGCTCTTATGAGGAAGTCTTATGAGGAAGTCAAACTATAAGCT 6448

QY 277 AAAAGATGAGGATATTTATATAGGTTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
Db 6447 TAAAGACGCTGATATTCATGAAGTCAAGGGTGGTATATCATCAAGGTGATGAAA 6388

QY 337 ATACTATGTTTACCTTAAGATGCTGCCAGCGGATACGTCGTCACAAAGAGGAAT 396
Db 6387 ATATATGTTTACCTTAAGATGCTGCCAGCGGATACGTCGTCGATGATATTCGAACTAAGATGAAT 6328

QY 397 CAATCGACAAACAAAGAGATAGTCAACATCGTGAAGGTGGAATCTCAAGAAACGATGG 456
Db 6327 CAATCGTCAAAACAAAGAGATAGTCAAGATGAAGTGAAGGTTA-----ACTC 6280

QY 457 TGCTGTTGCCCTTGGCAGCTTCGCAAGGACGCTACTACTACAGATGATGTTATATCTTAA 516
```

```
Db 6279 TAAATGTTGCTAGCAAGGTCTCAGGAGCGATATACGACAAATGATGTTATCTTTAA 6220
QY 517 TGCTTCTGATATCATAGAGATCTGGTGTATGCTTATATCGTTCTCATGGAGATCATTA 576
Db 6219 TCCAGCTGATATTTATCAAGATACGGGTATGCTTATATCGTTCTCATGGAGGTACATA 6160
QY 577 CCATTACATTCCTAAGAATGAGTTATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCCT 636
Db 6159 TCATTACATTCCTAAGAATGAGTTATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCCT 6100
QY 637 ATCTGTCGAGGAAATCTGTCAAAATCAAGAACCTATCGCCGACAAATAGCGATACAC 636
Db 6099 GGCTGGAAGAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGCTAGTAGCAAA--- 6043
QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATACAAATACTAACCAAG 756
Db 6042 -----TACACGCAATCTGT 6028
QY 757 CAACAAACAGCAACACTAACAAGTCAAGCAAGTCAAAAGTAAATGACATTTGATGTTGAA 816
Db 6027 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAATCTGAAAATCTCCAGAGTCTTTTGA 5968
QY 817 ACAGCTCTCAAAACTGCTTTGAGTCAACGACATGAGATCTGATGGCCTTCTTTGA 876
Db 5967 GGAATCTATGATTACCTAGCGCCCAACGTTACAGTGAATCAGATGGCCTGCTTTGA 5908
QY 877 TCCAGCACAAATCACAAGTCGAACAGCTAGAGGTGTTGCAAGTCCACACGAGATCATTA 936
Db 5907 CCTGCTAGATTTATCAGTGTACCAAAATGAGTTGCGATTCGCGATCGGACCATTA 5848
QY 937 CCACTTCTATCCTTACTCTCAAAATGCTGAATTTGGAAGAACGAATCGCTGATTTATTC 996
Db 5847 CCACTTTATCTTACAGCAAGCTTTCTGCTTTAGAAAGAAAGATTGCCAGAAATGGTGCC 5788
QY 997 CCTTCGT 1003
Db 5787 TATCAGT 5781

RESULT 14
US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Hp Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
```

:	TELEFAX:	(301) 309-8512
:	INFORMATION FOR SEQ ID NO:	355:
:	SEQUENCE CHARACTERISTICS:	
:	LENGTH:	973 base pairs
:	TYPE:	nucleic acid
:	STRANDEDNESS:	double
:	TOPOLOGY:	linear
:	US-08-961-527-355	

Query Match	16.0%;	Score	381.4;	DB	4;	Length	973;	
Best Local Similarity	77.5%;	Pred.	No. 8.7e-96;					
Matches	492;	Conservative	0;	Mismatches	131;	Indels	12; Gaps	2;

QY	1	TTCCTTACGAGTTGGACGTGATCAAGCTAGAACGGTTAAAGGAAAA---TAATCGTGTTTC	57
Dd	157	TTCCTATGAGCTTGGACGTTACCAGCTGGTCAGGATAAGAAGAGTCTAATCGAGTTGC	216
QY	58	CTATATAGATCGAAAAACAAGCGACGCCAAAAAACCGGAGAAATTGACTCCTCATGAGGTTAG	117
Dd	217	TTATATAGATCGTGATCAGGCTGGTCCAAAAGCGAGAAAACTTGACACCGATGAAGTCAG	276
QY	118	CAAGCGTGAAGGAATCAATGCTGAGCAAATCGTCATCAAGNATACACACCAAGGCTATGT	177
Dd	277	TAAAGGAGGGGATCAAGCCGAAACAAATGTTTATCAAGATTACGGATCAAGGTTATGT	336
QY	178	CAC TTCACATGGCGACCACTATCATATTATTACAATGGTAAGGTTCCCTTATCACGCTATCAT	237
Dd	337	GACCTCTCATGGAGACCAATTATCATTTACTATAATGCGAAGTTCCCTTATGATGCCATCAT	396
QY	238	CAGTGAAGMATTACTCATGAAGATCCAAACTATTAAGCTAAAAGATGAGGATATTGTTAA	297
Dd	397	CAGTGAAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGGATTCGACACATTGTCAA	456
QY	298	TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTAAAGGA	357
Dd	457	TGAATCAAGGGTGGTTATGTGCTATTAAGGTAAACGGTAAATATACTATGTTTACCTTAAAGGA	516
QY	358	TGCTGCCACCGCGATACGTCGTCGTAACAAAGAGGAAATCAATCGACAAAAACAAGACA	417
Dd	517	TGCAGCTCATCGCGATAATATTCCGACAAAAGAAGAGATTAAAACGTCAGAAGCAGGAAACG	576
QY	418	TAGTCAACATCGTGAAGGTGAACTCCAAGAAACGATGGTGCTGTTGSCCTTGACACGTTCC	477
Dd	577	CAGTCATAATCAT-----AACTCAAGAGCAGATTAATGCTGTGCTGCACCGCAGGC	627
QY	478	GCAAGGACGCTACTACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGGA	537
Dd	628	CCAAGGACGTTATACACCGGATATGGGTATATCTTCAATGCATCTCATATCATTCAGGA	687
QY	538	TACTGGTGTATGCTTATATCGTTCTCATGAGAGATCATTACCATTTACATTCCTTAAGATGA	597
Dd	688	CACGGGTGATGCTTATATCGTTCTCTCACGCGACCATTAACCATTAATTCCTTAAGATGA	747
QY	598	GTTATACGTAGCGAGTTGGCTGCTGTCAGAAGCCT	632
Dd	748	GTTATACGTAGCGAGTTAGCTGCTGAGAAGCCT	782

```

RESULT 15
US-08-961-527-258
; Sequence 258, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 258:
SEQUENCE CHARACTERISTICS:
LENGTH: 1684 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-258

Query Match	10.2%	Score 243.4	DB 4	Length 1684
Best Local Similarity	60.3%	Pred. No. 2.1e-57		
Matches 493	Conservative 0	Mismatches 236	Indels 88	Gaps 2
QY	1629	ATGCATATGTAACGCGCTCATATGGCCCATAGTCACATGGATTGGAAGAAGATAGCCTTTCG	1688	
Db	1			
QY	1689	ATAAGGAAAAAGTTGCAGCTCAAGCGCTATACTAAAGAAAAAGGTATCCTACCTCCATCTC	1748	
Db	61			
QY	1749	CAGACCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTTACAATCGTG	1808	
Db	120	CAGACCATCAGGATTCAGGAAATCTCAGGCAAAAGGAGCAGAAGCTATCTCAACACCGCG	179	
QY	1809	TGAAGCGGAAAAACGAATTCACACTCGTTCGACTCCATATATGGTTGAGCATACAGTTG	1868	
Db	180	TGAAGCAGCTAAGAAGTGCCACTTGATCGTATGCTTACAATCTTCAATATCTGTAG	239	
QY	1869	AGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCATTTACCATAATATTTAAATTTG	1928	
Db	240	AAGTCAAAAACGGTAGTTTAAATCATACCTCATTTATGACCATTACCATAACATCAAAATTG	239	
QY	1929	CTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGGAAGATTTGTTG	1988	
Db	300	AGTGGTTTGACGAAGCGCTTTTATGAGCGCACTAAGGGGTATCTCTTGAGGATCTTTTGG	359	
QY	1989	CGACGATTAACTACTACGTAGAACACCCCTGACGACGTCACATCTCATATGATGGATGGG	2048	
Db	360	CGACTGTCAGTACTATGTGGAACATCCAAACGACGTCGGCATTCAGATAATGGTTTG	419	
QY	2049	GCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAATAAGAACT	2108	
Db	420	GTAACGCTAGCGACCATGTTCAAGAGAAACAAAAATGGTCAAGCTGATACCAATCAACCG	479	
QY	2109	TCAAAGCGGATGA-----	2121	
Db	480	AAAAACCAAGCGGAGGAGAAACCTCAGACAGAAAAAACCCTGAGGAAGAAACCCCTCGAGAAG	539	
QY	2122	-----	-----	-----
Db	540	AGAAACCGCAAGCGGAGAAACAGAGTCTCAAAACCAACAGAGGAACCGAGAAATCAC	599	
QY	2142	CTTGCTGACCGAAGTTCCTCAAGTAGAGACTGAAAAAGTAGAAGGCCCAACTCAAGAAG	2201	

Db 600 CAGAGGAATCAGAAAGACCTCAGGTCGAGACTGAAAAGGTTGAGAAAACTCAGAGAGG 659
 QY 2202 CAGAAGTTTGGTGGAAAGTAACGATTCAGTCTGAAAGCCATGCAACAGAACTC 2261
 Db 660 CTGAAGATTTACTTGGAAAAATCCAGGATCCAAATTATCAAGTCCAATGCGCAAGAGACTC 719
 QY 2262 TAGCTGGTTTACGAAATAAATTTGACTCTTCCAAATTATGGATAACAATAGTATCATGSCAG 2321
 Db 720 TCACAGATTAAAAATAAATTTACTATTGCGCCAGGACCAACAATACTATTATGGCAG 779
 QY 2322 AAGCAGAAAAATTTACTTGGCTTTGTTAAAGGAAAGTAA 2358
 Db 780 AAGCTGAAACACTATTGCTTTTATTAAAGGAGAGTAA 816

RESULT 16
 US-08-743-637B-34/c
 ; Sequence 34, Application US/08743637B
 ; Patent No. 5994066
 ; GENERAL INFORMATION:
 ; APPLICANT: BERGERON, Michel G.
 ; APPLICANT: PICARD, Francois J.
 ; APPLICANT: OUELLETTE, Marc
 ; APPLICANT: ROY, Paul H.
 ; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
 ; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
 ; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
 ; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
 ; NUMBER OF SEQUENCES: 273
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: QUARLES & BRADY
 ; STREET: 411 EAST WISCONSIN AVENUE
 ; CITY: MILWAUKEE
 ; STATE: WISCONSIN
 ; COUNTRY: USA
 ; ZIP: 53202-4497
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/743,637B
 ; FILING DATE: 04-NOV-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/526,840
 ; FILING DATE: 11-SEP-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BAKER, Jean C.
 ; REGISTRATION NUMBER: 35,433
 ; REFERENCE/DOCKET NUMBER: 850586.90012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (414) 277-5000
 ; TELEFAX: (414) 277-5591
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 841 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; US-08-743-637B-34

Query Match 4.6%; Score 109; DB 2; Length 841;
 Best Local Similarity 65.4%; Pred. No. 2.7e-20;
 Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;
 QY 2090 GAAGATCCAAATAAGAACTTCAACGGATGAGAGCCAGTAGAGGAAACACCTGCTGAG 2149
 Db 633 GAGTCTCCAAAACCAACAGAGGAAACCAAGAAAGAAATCACCAGAAATCACCAGAGAA 574

QY 2150 CCAGAAGTCCCTCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAAAGAGCAGAGTT 2209
 Db 573 TCAGAAGAACCCTCAGGTCGAGACTGAAAAGGTTAAGAA--AACTGAGAGAGGCTGAAGAT 517
 QY 2210 TTGCTTGGAAAGTAACGATTCAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGT 2269
 Db 516 TTACTTTGAAAAATCCAGAAATCCAAATTTATCAAGTCCAAATGCCAAAGAGACTCTCACAGGA 457
 QY 2270 TTACGAAATAAATTTGACTCTTCAAAATTTATGATAACAATAGTATCATGCGAGAGCAGAA 2329
 Db 456 TTAATAAAATAAATTTACTATTGCGCCAGGACCAACAATACTATTATGGCAGAGCTGAA 397
 QY 2330 AAATTTACTTGGCTTTGTTAAAGGAAAGTAA 2358
 Db 396 AAATTTACTTGGCTTTTATTAAAGGAGAGTAA 368

RESULT 17
 US-08-526-840B-34/c
 ; Sequence 34, Application US/08526840B
 ; Patent No. 6001564
 ; GENERAL INFORMATION:
 ; APPLICANT: BERGERON, Michel G.
 ; APPLICANT: OUELLETTE, Marc
 ; APPLICANT: ROY, Paul H.
 ; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
 ; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
 ; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
 ; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
 ; NUMBER OF SEQUENCES: 177
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: QUARLES & BRADY
 ; STREET: 411 East Wisconsin Avenue
 ; CITY: Milwaukee
 ; STATE: Wisconsin
 ; COUNTRY: USA
 ; ZIP: 53202-4497
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/526,840B
 ; FILING DATE: 11-SEP-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/304,732
 ; FILING DATE: 12-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BAKER, Jean C.
 ; REGISTRATION NUMBER: 35,433
 ; REFERENCE/DOCKET NUMBER: 850586.90012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (414) 277-5000
 ; TELEFAX: (414) 277-5591
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 841 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; US-08-526-840B-34

Query Match 4.6%; Score 109; DB 3; Length 841;
 Best Local Similarity 65.4%; Pred. No. 2.7e-20;
 Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;
 QY 2090 GAAGATCCAAATAAGAACTTCAACGGATGAGAGCCAGTAGAGGAAACACCTGCTGAG 2149

```

Db 633 GAGTCTCCAAAACCAACAGAGGACCAAGAGAAATCACCAGAGAAATCACCAGAGAA 574
Qy 2150 CCAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAGAGAGAGTT 2209
Db 573 TCAGAGAACTCAGGTCGAGACTGAAAGGTTAAGAA--AACTGAGAGAGGCTGAAGAT 517
Qy 2210 TTGCTTCGGAAGTAACGGATTCTAGTCTCAAGCCCAATCAACAGAGAACTCTAGCTGGT 2269
Db 516 TTACTTGGAATAATCAGAAATCAATATCAAGTCCATGCCAAAGAGACTCTCAGGAA 457
Qy 2270 TTACGAAATAATTTGACTCTTCAAAATATGAGATAACAATAGTAGTGGCAGAGAGAGAA 2329
Db 456 TTAAAAATAATTTACTATTGTCACCCAGGCAACAATACTATTATGGCAGAGAGCTGAA 397
Qy 2330 AAATTACTTCGGTGTGTAAGAGAGTAA 2358
Db 396 AAATATTGGCTTTATTAAAGGAGAGTAA 368

```

```

RESULT 18
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

```

```

Query Match 2.4%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 3.4%; Pred. No. 1.7e-05;
Matches 13; Conservative 224; Mismatches 150; Indels 0; Gaps 0;

```

```

Qy 1974 TGGAAGATTGTTTGGCAGCAATTAAGTACTAGTAGAAACCCCTGACGAAACGTCCACATT 2033
Db 1450 TAGAAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
Qy 2034 CTAATGATGCGGCAATGCCAGTCAGCATGTGTTAGGCAAGAAAGACACAGTGAAG 2093
Db 1390 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1331
Qy 2094 ATCAAAATAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGAGAAACACCTGCTGAGCCAG 2153
Db 1330 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271
Qy 2154 AAGTCCTCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAGAGAGAGAGTTTTCG 2213
Db 1270 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1211
Qy 2214 TTCCGAAAGTAACGCAATCTAGTCTCAAGCCCAATCAACAGAGAACTCTAGCTGGTTTAC 2273
Db 1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151
Qy 2274 GAAATAATTGACTCTTCAAAATATGATAACAATATGATGCGAGAGAGAGAGAGAGAG 2333
Db 1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091
Qy 2334 TACTTCGCTGTGTTAAAGGAAGTAATC 2360
Db 1090 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1064

```

```

RESULT 19
US-09-134-001C-2243
; Sequence 2243, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2243
; LENGTH: 11091
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2243

```

```

Query Match 2.2%; Score 51.6; DB 4; Length 11091;
Best Local Similarity 47.3%; Pred. No. 0.00088;
Matches 187; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

```

```

Qy 1096 AAATCTTAAATAGACTCAAAATCTCTTTGTTAGTCAGTCGGTACGAAAGTTGGGA 1155
Db 8481 AAAAGCCAAAATGATGTAATCAATCTCAAACTAATCAGCAAGTTGAAATGCTGAGCA 8540
Qy 1156 AGGATATGTTATTCGAAGAAAAGGGCATCTCTGTTATGCTTTGCGAAA--GATTACC 1212
Db 8541 AAATAGTTTAGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 8600
Qy 1213 ATCTGAAACTGTAAAAAATCTTGAAGCAAGTTTATCAAAACAGAGAGTGTTCACACAC 1272
Db 8601 AGCTGAAATTTTAAAGCGCAACAAAACAAATTCATGAAATAGAGCAAGATTTAGTGC 8660
Qy 1273 TTAACTGCTAAAAAGAAAATGTTGCTCTCTGTCGACCAAGATTTTATGATAAGACATA 1332
Db 8661 TACAAGAGAGAAAAGCAATGCTTTTACAACATTTAGATGAACAGGTTAAAGAAATCAT 8720
Qy 1333 TAATCTGTTAACTCAGGCTCATAAAGCCTGTTTGNAAATAAGGCTGTAATTTCTGATT 1392

```

```

Db      8721 TAATTTATTAATCAAGCTAATACAGATATAGTAGATATAGTCTAAACCTTCGGTT 8780
QY      1393 CCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGCATTAATAAGAAAAT 1452
Db      8781 GAATAACATAACTGAATACAGACAGACAGATATAATAAAAGAAAATGCTATATTAAAT 8840
QY      1453 GGTAGATGATTATTGGCATTCCTAGCACCAATTA 1487
Db      8841 ATATGATGTTTCAGATACCTCAAGAAAGCTATAATTA 8875

RESULT 20
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-2

Query Match      1.9%; Score 45.2; DB 1; Length 2277;
Best Local Similarity 29.7%; Pred. No. 0.022;
Matches 77; Conservative 51; Mismatches 131; Indels 0; Gaps 0;

QY      279 AAGATGAGGATATTGTTAATGAGTCAAGGTGGATATGTTATCAAGGTAGATGAAAAT 338
Db      671 ARGARGAGATGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARG 730
QY      339 ACTATGTTTACCTTAAGGATGCTGCCACGGGATAACGTCCTACAAAAGAGGAAATCA 398
Db      731 AYGGNGTNTTYGAYGARGARGARGARAAYATHGARWSNARGTNACNAARC 790
QY      399 ATCGACAAAACAAGACGATAGTCAACATCGTGAAGGTGGAATCTCCAAAGAACGATGTG 458
Db      791 CNGTNCARATHCARAARMGNCNGTNAARMGNCNGCNCNGCNAARWSNNGAYCAYW 850
QY      459 CTGTTGCCCTTGGCAGCTTCGCAAGACGCTATACAGATGATGTTTATCTTTAATG 518
Db      851 SNGARGARGAYWSNGAYTNGARGARWSNATHGAYGAYGNGARGARYTNGCNC 910
QY      519 CTTCGTGATCATAGAGGA 537
Db      911 ARWSNGAYACNWSNACNGA 929

US-09-765-271-55.rni

Db      911 ARWSNGAYACNWSNACNGA 929

RESULT 21
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-2

Query Match      1.9%; Score 45.2; DB 1; Length 2277;
Best Local Similarity 29.7%; Pred. No. 0.022;
Matches 77; Conservative 51; Mismatches 131; Indels 0; Gaps 0;

QY      279 AAGATGAGGATATTGTTAATGAGTCAAGGTGGATATGTTATCAAGGTAGATGAAAAT 338
Db      671 ARGARGAGATGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARG 730
QY      339 ACTATGTTTACCTTAAGGATGCTGCCACGGGATAACGTCCTACAAAAGAGGAAATCA 398
Db      731 AYGGNGTNTTYGAYGARGARGARGARAAYATHGARWSNARGTNACNAARC 790
QY      399 ATCGACAAAACAAGACGATAGTCAACATCGTGAAGGTGGAATCTCCAAAGAACGATGTG 458
Db      791 CNGTNCARATHCARAARMGNCNGTNAARMGNCNGCNCNGCNAARWSNNGAYCAYW 850
QY      459 CTGTTGCCCTTGGCAGCTTCGCAAGACGCTATACAGATGATGTTTATCTTTAATG 518
Db      851 SNGARGARGAYWSNGAYTNGARGARWSNATHGAYGAYGNGARGARYTNGCNC 910
QY      519 CTTCGTGATCATAGAGGA 537
Db      911 ARWSNGAYACNWSNACNGA 929

US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:

```

```
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-098-487-2

Query Match
Best Local Similarity 1.9%; Score 45.2; DB 2; Length 2277;
Matches 77; Conservative 51; Mismatches 131; Indels 0; Gaps 0;

QY 279 AAGATGAGGATATTGTTAATGAGTCAAGGTCGATGTTATCAAGGTAGATGGAAT 338
Db 671 ARGARGARGATGARGARGARGARGARGARGARGARGARGARGARGARGARGARG 730
QY 339 ACTATGTTTACCTTAAGGATGTCGCCACCGGATACGTCGATCAAAAAGAGAAATCA 398
Db 731 AYGCGTNTTYGAYGARGAYGARGARGARGARGARGARGARGARGARGARGARGARG 790
QY 399 ATCGACAAAACAGAGCATAGTCAACATGCTGGAAGTGGAACTCCAAGAAACGATGGTG 458
Db 791 CNGTNCARATHCARARMGNCNGTNAARMGNCNCNGCNGCNGCNGCNGCNGCNGCNGC 850
QY 459 CTGTTGCTTGGCAGCTTCGCAAGGAGCGTATACATGATGATGATGATGATGATG 518
Db 851 SNGARGARGAYNSGAYNYNGARGARGARGARGARGARGARGARGARGARGARGARG 910
QY 519 CTCTGATATCATAGAGGA 537
Db 911 ARWSNGAYACNWSNACNGA 929

RESULT 23
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335

; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidlund, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3945
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-6

Query Match
Best Local Similarity 1.8%; Score 42.4; DB 4; Length 3945;
Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1695 AAAAAGTTGCAGCTCAAGCTATATAAGAAAAAGGATATCTCATCTCATCTCCAGAG 1754
Db 365 ATAAAGTTGAAGTTTCAACTGCCAATCAGATGAGCAAGCTTCAACAAATCTACGAATG 424
```

```
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
; US-09-621-976-2813

Query Match
Best Local Similarity 1.8%; Score 44; DB 4; Length 832;
Matches 29; Conservative 160; Mismatches 135; Indels 0; Gaps 0;

QY 97 TTGACTCCTGATGAGGTTAGCAAGCTGAGGAATCAATCTGAGCAAAATCGTCATCAA 156
Db 348 YTTTMMWKKARRWYWKSTYACASRYKTYGWWYMMWYMMWYMMWYMMWYMMWYMMW 289
QY 157 GATAACAGACCAAGGCTATGCTTCACTTCCATGCGGACCATCATATTATTAATGTA 216
Db 288 YRGRCAYTMRGRMWSYAWGKWSMRSAKMSCTMYKKGSTYMTKCTCATWCYWM 229
QY 217 GGTTCCTTATGACGCTATCATCTAGTGAAGATTAATCAAGATCAAACTAAGCT 276
Db 228 KYKRWMSKTCWSRGYNTSTSYWASWYTWCMWGRWWSYTWYMAWGW 169
QY 277 AAAAGATGAGGATATTGTTAATGAGTCAAGGTCGATGTTATCAAGGTAGATGAAA 336
Db 168 WRYATTWRRAMWMAAWTMMWYMAWCSRGAAMYRRTMMWGYWYWKSYRTR 109
QY 337 ATACTATGTTTACCTTAAGGATGCTGCCACGGGATAAGCTCGGTACAAAAGAGAAAT 396
Db 108 CAWAYANKTKRSYTCWCRWKRCMMMMMAAYGKTMMAACWKTTRYWRWAWMWRMW 49
QY 397 CAATCGACAAAAACAAGACGATAG 420
Db 48 WTMMWYWRAMKRRWRKWR 25
```

```
RESULT 24
US-09-200-650E-6
; Sequence 6, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidlund, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3945
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-6

Query Match
Best Local Similarity 1.8%; Score 42.4; DB 4; Length 3945;
Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1695 AAAAAGTTGCAGCTCAAGCTATATAAGAAAAAGGATATCTCATCTCATCTCCAGAG 1754
Db 365 ATAAAGTTGAAGTTTCAACTGCCAATCAGATGAGCAAGCTTCAACAAATCTACGAATG 424
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 15:51:17 ; Search time 1122 Seconds

(without alignments)
10794.292 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTCGGACTGT.....TAAGTAAGGAATAAATAC 2389

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
18:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
19:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2388	100.0	2389	9	US-09-765-272-55
2	2388	100.0	2451	13	US-10-412-862-9
3	2388	100.0	2451	16	US-10-412-850-9
4	2388	100.0	2451	16	US-10-387-783-9
5	2388	100.0	8195	13	US-10-158-844-94
6	1374.2	57.5	2523	10	US-09-884-465A-3
7	1374.2	57.5	2647	10	US-09-769-787-206
8	1003.8	42.0	2481	10	US-09-884-465A-4
9	991	41.5	2531	13	US-10-412-862-11
10	991	41.5	2531	16	US-10-412-850-11
11	991	41.5	2531	16	US-10-387-783-11
12	990	41.4	2531	13	US-10-412-862-5
13	990	41.4	2531	16	US-10-412-850-5
14	990	41.4	2531	16	US-10-387-783-5
15	990	41.4	2531	16	US-10-387-783-5
16	990	41.4	2531	16	US-10-387-783-5
17	990	41.4	2531	16	US-10-387-783-5
18	990	41.4	2531	16	US-10-387-783-5
19	990	41.4	2531	16	US-10-387-783-5
20	990	41.4	2531	16	US-10-387-783-5
21	990	41.4	2531	16	US-10-387-783-5
22	990	41.4	2531	16	US-10-387-783-5
23	990	41.4	2531	16	US-10-387-783-5
24	990	41.4	2531	16	US-10-387-783-5
25	990	41.4	2531	16	US-10-387-783-5
26	990	41.4	2531	16	US-10-387-783-5
27	990	41.4	2531	16	US-10-387-783-5
28	990	41.4	2531	16	US-10-387-783-5
29	990	41.4	2531	16	US-10-387-783-5
30	990	41.4	2531	16	US-10-387-783-5
31	990	41.4	2531	16	US-10-387-783-5
32	990	41.4	2531	16	US-10-387-783-5
33	990	41.4	2531	16	US-10-387-783-5
34	990	41.4	2531	16	US-10-387-783-5
35	990	41.4	2531	16	US-10-387-783-5
36	990	41.4	2531	16	US-10-387-783-5
37	990	41.4	2531	16	US-10-387-783-5
38	990	41.4	2531	16	US-10-387-783-5
39	990	41.4	2531	16	US-10-387-783-5
40	990	41.4	2531	16	US-10-387-783-5
41	990	41.4	2531	16	US-10-387-783-5
42	990	41.4	2531	16	US-10-387-783-5
43	990	41.4	2531	16	US-10-387-783-5
44	990	41.4	2531	16	US-10-387-783-5
45	990	41.4	2531	16	US-10-387-783-5
46	990	41.4	2531	16	US-10-387-783-5
47	990	41.4	2531	16	US-10-387-783-5
48	990	41.4	2531	16	US-10-387-783-5
49	990	41.4	2531	16	US-10-387-783-5
50	990	41.4	2531	16	US-10-387-783-5
51	990	41.4	2531	16	US-10-387-783-5
52	990	41.4	2531	16	US-10-387-783-5
53	990	41.4	2531	16	US-10-387-783-5
54	990	41.4	2531	16	US-10-387-783-5
55	990	41.4	2531	16	US-10-387-783-5
56	990	41.4	2531	16	US-10-387-783-5
57	990	41.4	2531	16	US-10-387-783-5
58	990	41.4	2531	16	US-10-387-783-5
59	990	41.4	2531	16	US-10-387-783-5
60	990	41.4	2531	16	US-10-387-783-5
61	990	41.4	2531	16	US-10-387-783-5
62	990	41.4	2531	16	US-10-387-783-5
63	990	41.4	2531	16	US-10-387-783-5
64	990	41.4	2531	16	US-10-387-783-5
65	990	41.4	2531	16	US-10-387-783-5
66	990	41.4	2531	16	US-10-387-783-5
67	990	41.4	2531	16	US-10-387-783-5
68	990	41.4	2531	16	US-10-387-783-5
69	990	41.4	2531	16	US-10-387-783-5
70	990	41.4	2531	16	US-10-387-783-5
71	990	41.4	2531	16	US-10-387-783-5
72	990	41.4	2531	16	US-10-387-783-5
73	990	41.4	2531	16	US-10-387-783-5
74	990	41.4	2531	16	US-10-387-783-5
75	990	41.4	2531	16	US-10-387-783-5
76	990	41.4	2531	16	US-10-387-783-5
77	990	41.4	2531	16	US-10-387-783-5
78	990	41.4	2531	16	US-10-387-783-5
79	990	41.4	2531	16	US-10-387-783-5
80	990	41.4	2531	16	US-10-387-783-5
81	990	41.4	2531	16	US-10-387-783-5
82	990	41.4	2531	16	US-10-387-783-5
83	990	41.4	2531	16	US-10-387-783-5
84	990	41.4	2531	16	US-10-387-783-5
85	990	41.4	2531	16	US-10-387-783-5
86	990	41.4	2531	16	US-10-387-783-5
87	990	41.4	2531	16	US-10-387-783-5
88	990	41.4	2531	16	US-10-387-783-5
89	990	41.4	2531	16	US-10-387-783-5
90	990	41.4	2531	16	US-10-387-783-5

Sequence 65, Appli
Sequence 5, Appli
Sequence 243, App
Sequence 181, App
Sequence 246, App
Sequence 23, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 192, App
Sequence 355, App
Sequence 9, Appli
Sequence 258, App
Sequence 23, Appli
Sequence 17, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 257, App
Sequence 10, Appli
Sequence 6, Appli
Sequence 71, Appli
Sequence 34, Appli
Sequence 1825, Ap
Sequence 35051, A
Sequence 2179, Ap
Sequence 21533, A
Sequence 4012, Ap
Sequence 20619, A
Sequence 3, Appli
Sequence 3, Appli
Sequence 18355, A
Sequence 1597, Ap
Sequence 2, Appli
Sequence 35506, A
Sequence 19608, A
Sequence 2825, Ap
Sequence 34866, A
Sequence 20406, A
Sequence 463, App
Sequence 1855, Ap
Sequence 26267, A
Sequence 71, Appli
Sequence 19488, A
Sequence 2772, Ap
Sequence 4843, Ap
Sequence 9039, Ap
Sequence 454, App
Sequence 454, App
Sequence 269927,
Sequence 269927,
Sequence 19262, A
Sequence 2534, Ap
Sequence 6282, Ap
Sequence 1913, Ap
Sequence 1569, Ap
Sequence 77, Appli
Sequence 5350, Ap
Sequence 24901, A
Sequence 1, Appli
Sequence 378, App
Sequence 14986, A
Sequence 1284, Ap
Sequence 233, App
Sequence 313, App
Sequence 50, Appli
Sequence 1928, Ap
Sequence 1928, Ap
Sequence 8, Appli

c 88 41.6 1.7 3697 13 US-10-363-616-178 Sequence 178, App
89 41.6 1.7 60787 13 US-10-087-192-1774 Sequence 1774, App
90 41.4 1.7 2721 13 US-10-282-122A-40640 Sequence 40640, A

ALIGNMENTS

RESULT 1
US-09-765-272-55
; Sequence 55, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-765-272-55

Query Match 100.0%; Score 2388; DB 9; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60
DB 1 TTCCTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60
QY 61 TATAGATCGAAAACAAAGCGACGCAAAACCGAGAAATTGACTCCTGATGAGGTTAGCAA 120
DB 61 TATAGATCGAAAACAAAGCGACGCAAAACCGAGAAATTGACTCCTGATGAGGTTAGCAA 120
QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAGGCTATGTCA 180
DB 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAGGCTATGTCA 180
QY 181 TTCATATGCGACCACTATCATTTATCAATGGTAAGGTTTCCTATGACGCTATCATCAG 240
DB 181 TTCATATGCGACCACTATCATTTATCAATGGTAAGGTTTCCTATGACGCTATCATCAG 240
QY 241 TGAAGATTACTCATGAAGATCCAACTAATAGCTAAAGATGAGGATTTGTTAATGA 300

241 TGAAGATTACTCATGAAAAGATCCAAACTATATAAGCTAAAGATGAGGATTTGTTAATGA 300
QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGC 360
DB 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGC 360
QY 361 TGCCACGCGGATACGTCCTGTAACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
DB 361 TGCCACGCGGATACGTCCTGTAACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGCTTGCCTTGCACGTTCCGCA 480
DB 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGCTTGCCTTGCACGTTCCGCA 480
QY 481 AGGACGCTATACCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
DB 481 AGGACGCTATACCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY 541 TGGTGATGCTTATATGCTTCTCATGGAGATCATATTACCATTAATTCCTTAAGATGAGTT 600
DB 541 TGGTGATGCTTATATGCTTCTCATGGAGATCATATTACCATTAATTCCTTAAGATGAGTT 600
QY 601 ATCAGCTAGCGAGTTGCTGCTGCAGAGGCTTCTTATCTGCTCGAGGAAATCTGTCAAA 660
DB 601 ATCAGCTAGCGAGTTGCTGCTGCAGAGGCTTCTTATCTGCTCGAGGAAATCTGTCAAA 660
QY 661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACATTCAGAAACAAACTGGGTACCTTC 720
DB 661 TTCAAGAACCTATCGCCGACAAAATAGCGATAACACATTCAGAAACAAACTGGGTACCTTC 720
QY 721 TGTAAGCAATCCAGGAACTACAAATATTAACAAGCAACAAAGCAACAACTAACAGTCA 780
DB 721 TGTAAGCAATCCAGGAACTACAAATATTAACAAGCAACAAAGCAACAACTAACAGTCA 780
QY 781 AGCAAGTCAAGATTAATGACATTTGATGCTCTTCAAAACAGCTCTACAACTGCTTTGAG 840
DB 781 AGCAAGTCAAGATTAATGACATTTGATGCTCTTCAAAACAGCTCTACAACTGCTTTGAG 840
QY 841 TCACAGACATGTAGAATCTGATGGCTTTGCTTTGATCCAGCACAAATCACAAGTCGAAC 900
DB 841 TCACAGACATGTAGAATCTGATGGCTTTGCTTTGATCCAGCACAAATCACAAGTCGAAC 900
QY 901 AGCTAGAGGTTTGCAGTGCCACACGGAGATCATATCCACTTCACTTCTACTCTCAAT 960
DB 901 AGCTAGAGGTTTGCAGTGCCACACGGAGATCATATCCACTTCACTTCTACTCTCAAT 960
QY 961 GTCTGAATTTGGAAGCAAGATCGCTGCTGTTATTTTCCCTTCTGTTATCGTTCAAAACCAT 1020
DB 961 GTCTGAATTTGGAAGCAAGATCGCTGCTGTTATTTTCCCTTCTGTTATCGTTCAAAACCAT 1020
QY 1021 GGTACCAAGATTCAAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACTAGTCCAGG 1080
DB 1021 GGTACCAAGATTCAAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACTAGTCCAGG 1080
QY 1081 CCGCAACCTCGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTGTAGTCTAGCTGGT 1140
DB 1081 CCGCAACCTCGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTGTAGTCTAGCTGGT 1140
QY 1141 ACGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAAGGGCATCTCTCGTTATGTTTGC 1200
DB 1141 ACGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAAGGGCATCTCTCGTTATGTTTGC 1200
QY 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTTTGAAAAGCAAGTTATCAAAACAAGAGAG 1260
DB 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTTTGAAAAGCAAGTTATCAAAACAAGAGAG 1260
QY 1261 TGTTTCAACACTTTAACTGCTTAAAGAAAATGTTGCTCTCGTCCGACCAAGAAATTTTA 1320
DB 1261 TGTTTCAACACTTTAACTGCTTAAAGAAAATGTTGCTCTCGTCCGACCAAGAAATTTTA 1320
QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGGTGG 1380
DB 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGGTGG 1380

1381 TAAATCTGATTTCCAGCGCTTAGACAAATATTAGAACGCTTGAATGATGATGACGTAA 1440
1381 TAAATCTGATTTCCAGCGCTTAGACAAATATTAGAACGCTTGAATGATGATGACGTAA 1440
1441 TAAAGAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCACCATTACCCATCCAGAGCG 1500
1441 TAAAGAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCACCATTACCCATCCAGAGCG 1500
1501 ACTTGGCAACCAATTTCTCAATTTAGTATACCTGAAGACCAAGTTTGGTATGCTCAATT 1560
1501 ACTTGGCAACCAATTTCTCAATTTAGTATACCTGAAGACCAAGTTTGGTATGCTCAATT 1560
1561 AGCTGATAAGTATACCAACGTCAGATGTTTACATTTTGTGAAACATCATATTAATCACTGA 1620
1561 AGCTGATAAGTATACCAACGTCAGATGTTTGTGAAACATCATATTAATCACTGA 1620
1621 TGAAGGAGTGCATATGTAACGCTCATATGGGCCATAGTCACCTGGATTGGAAAAGATAG 1680
1621 TGAAGGAGTGCATATGTAACGCTCATATGGGCCATAGTCACCTGGATTGGAAAAGATAG 1680
1681 CCTTTCTGATAAGGAAAAGTTTCAGCTCAAGCTTATACCTAAGAAAAGGTTATCTTACC 1740
1681 CCTTTCTGATAAGGAAAAGTTTCAGCTCAAGCTTATACCTAAGAAAAGGTTATCTTACC 1740
1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTTA 1800
1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTTA 1800
1801 CAATCTGTGAAAGGGGAAAAAGAAATTCCTCGTTCCGACTTCCATATATGTTGAGCA 1860
1801 CAATCTGTGAAAGGGGAAAAAGAAATTCCTCGTTCCGACTTCCATATATGTTGAGCA 1860
1861 TACAGTTGAGTTTAAAGCGTAATTTGATTTATTCCTCATAGGATCATTAACATATAT 1920
1861 TACAGTTGAGTTTAAAGCGTAATTTGATTTATTCCTCATAGGATCATTAACATATAT 1920
1921 TAAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1921 TAAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1981 TTTGTTTGGACGATTAAGTATACCTAGAACACCTGACGAACTCCACATTTCTAATGA 2040
2041 TGGATGGGCAATGCCAGTGAGCATGTTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
2101 TGAAGAACTTTCAAAGCGGATGAAGGCCAGTAGAGGAAACCTGCTGAGCCAGAGTCCC 2160
2101 TGAAGAACTTTCAAAGCGGATGAAGGCCAGTAGAGGAAACCTGCTGAGCCAGAGTCCC 2160
2161 TCAAGTAGAGACTGAAAAGTTAGAGGCCCACTCAAGAAAGCAGAGTTTTCCTTGGCAA 2220
2161 TCAAGTAGAGACTGAAAAGTTAGAGGCCCACTCAAGAAAGCAGAGTTTTCCTTGGCAA 2220
2221 AGTAACGGATTTCTAGTCTCAAGCCCAATGCAACAGAACTTAGCTGTTTACGAAATTA 2280
2221 AGTAACGGATTTCTAGTCTCAAGCCCAATGCAACAGAACTTAGCTGTTTACGAAATTA 2280
2281 TTTGACTCTTCAAAATTTATGGATAACAATAGTATCATGGCAGAGCAGAAAATTAATTCTGC 2340
2281 TTTGACTCTTCAAAATTTATGGATAACAATAGTATCATGGCAGAGCAGAAAATTAATTCTGC 2340
2341 GTTGTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAATAAATTAAC 2389
2341 GTTGTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAATAAATTAAC 2389

RESULT 2

US-10-412-862-9

; Sequence 9, Application US/10412862

Publication No. US20040052781A1
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
TITLE OF INVENTION: Motifs
FILE REFERENCE: 469201-685
CURRENT APPLICATION NUMBER: US/10/412,862
CURRENT FILING DATE: 2003-04-14
PRIOR APPLICATION NUMBER: 09/468,656
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 9
LENGTH: 2451
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(2451)
OTHER INFORMATION: n = a, c, t or g
US-10-412-862-9

Query Match 100.0%; Score 2388; DB 13; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTCTAGGAGTTGGGACTGTATCAAGCTAGAACCGTTAAAGGAAAATAATCGTTTCTCTA 60
Db 60 TTTCTAGGAGTTGGGACTGTATCAAGCTAGAACCGTTAAAGGAAAATAATCGTTTCTCTA 119
QY 61 TATAGATGAAAACAAAGCGACGACCAAAACCGGAGAAATTTGACTCCCTGATGAGTTAGCAA 120
Db 120 TATAGATGAAAACAAAGCGACGACCAAAACCGGAGAAATTTGACTCCCTGATGAGTTAGCAA 179
QY 121 GCGTGAAGGAATCAATGCTGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 180 GCGTGAAGGAATCAATGCTGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 239
QY 181 TTCAATGCGGACCACTATCAATTTATCAATGGTAAAGTTCCCTATGACGCTATCATCAG 240
Db 240 TTCAATGCGGACCACTATCAATTTATCAATGGTAAAGTTCCCTATGACGCTATCATCAG 299
QY 241 TGAAGAATTTACTCATCAAGATCCAAACTATAAGCTTAAAGATGAGGATATGTTAATGA 300
Db 300 TGAAGAATTTACTCATCAAGATCCAAACTATAAGCTTAAAGATGAGGATATGTTAATGA 359
QY 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAAGGATGC 360
Db 360 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAAGGATGC 419
QY 361 TGCCCAACCGGATTAACGTCGGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 420
Db 420 TGCCCAACCGGATTAACGTCGGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 479
QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCACTGTCGA 480
Db 480 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCACTGTCGA 539
QY 481 AGGACGCTTACTACAGATGATGTTTATCTTTAAATGCTTCTGATATCATAGAGGATAC 540
Db 540 AGGACGCTTACTACAGATGATGTTTATCTTTAAATGCTTCTGATATCATAGAGGATAC 599
QY 541 TGGTGAATCTTATATCGTTTCTCATGAGATCAATTAACATTAATTCCTAAGAAATGAGTT 600
Db 600 TGGTGAATCTTATATCGTTTCTCATGAGATCAATTAACATTAATTCCTAAGAAATGAGTT 659
QY 601 ATCAGTAGCGAGTTGGCTGCTGAGAGCCCTTCTATCTGCTGAGGAAATCTGTCAA 660
Db 660 ATCAGTAGCGAGTTGGCTGCTGAGAGCCCTTCTATCTGCTGAGGAAATCTGTCAA 719

661 TTCAAGAACTATCGCGCAAAATAGCGATAACATCTCAAGAACTGGTACCTTC 720
 720 TTCAAGAACTATCGCGCAAAATAGCGATAACATCTCAAGAACTGGTACCTTC 779
 721 TGTAAAGCAATCCAGGAACTACAAATCTAACAAGCAACAGCAACAGCACTAACAGTCA 780
 780 TGTAAAGCAATCCAGGAACTACAAATCTAACAAGCAACAGCAACAGCACTAACAGTCA 839
 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTTCAAACTGGCTTTGAG 840
 840 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTTCAAACTGGCTTTGAG 899
 841 TCACAGCAATGTAGATCTGATGGCTTGTCTTTGATCCAGCAACATCAAGTCCGAAAC 900
 900 TCACAGCAATGTAGATCTGATGGCTTGTCTTTGATCCAGCAACATCAAGTCCGAAAC 959
 901 AGCTAGAGGTGTGCACTGGCCACACGAGATCATTAGCACTTCACTCCCTTACTCTCAAT 960
 960 AGCTAGAGGTGTGCACTGGCCACACGAGATCATTAGCACTTCACTCCCTTACTCTCAAT 1019
 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTCCTTTCCTTATCGTTCAAACTTTG 1020
 1020 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTCCTTTCCTTATCGTTCAAACTTTG 1079
 1021 GGTACCGATTTCAAGGCCAGAAACAAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 1080 GGTACCGATTTCAAGGCCAGAAACAAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139
 1081 CCCGCACTCCACCAATCTTAAATAGATCTCAAAATCTTCTTTGGTTAGTCACTGGT 1140
 1140 CCCGCACTCCACCAATCTTAAATAGATCTCAAAATCTTCTTTGGTTAGTCACTGGT 1199
 1141 ACAGAAAGTTGGGAAGGATATGATTCGAAGAAAGGGGATCTCTCGTTATGTTCTTTC 1200
 1200 ACAGAAAGTTGGGAAGGATATGATTCGAAGAAAGGGGATCTCTCGTTATGTTCTTTC 1259
 1201 GAAAGATTTACCACTGAACTGTTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG 1260
 1260 GAAAGATTTACCACTGAACTGTTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG 1319
 1261 TGTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTCTCTCGTGACCAAGATTTTA 1320
 1320 TGTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTCTCTCGTGACCAAGATTTTA 1379
 1321 TGATAAGCATATAATCTGTTAACTGAGCTCATAAAGCTTGTGTTGAAATAGGGTCG 1380
 1380 TGATAAGCATATAATCTGTTAACTGAGCTCATAAAGCTTGTGTTGAAATAGGGTCG 1439
 1381 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAACGCTTGAATGATGATCACTAA 1440
 1440 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAACGCTTGAATGATGATCACTAA 1499
 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATCTCTAGCACCAATTAACCATCCAGAGCG 1500
 1500 TAAAGAAATTTGGTAGATGATTTATTTGGCAATCTCTAGCACCAATTAACCATCCAGAGCG 1559
 1501 ACTTGGCAACCAATCTCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 1560 ACTTGGCAACCAATCTCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1619
 1561 ACTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 1620 ACTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679
 1621 TGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
 1680 TGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739
 1681 CCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
 1740 CCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1799

1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGGACAGCTATTTA 1800
 1800 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGGACAGCTATTTA 1859
 1801 CAATCTGTGTGAAGGGGAAAGCAATTCACCTCGTTCCAGTTCCTCATATATATGTTGAGCA 1860
 1860 CAATCTGTGTGAAGGGGAAAGCAATTCACCTCGTTCCAGTTCCTCATATATATGTTGAGCA 1919
 1861 TACAGTTGAGTTTAAAAACGGTAAATTTGATTTATTCCTCATAAAGGATCATTAACATAAT 1920
 1920 TACAGTTGAGTTTAAAAACGGTAAATTTGATTTATTCCTCATAAAGGATCATTAACATAAT 1979
 1921 TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTATACCTTTGGAGA 1980
 1980 TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTATACCTTTGGAGA 2039
 1981 TTTGTTTGGCAACCATTAAGTACTACGTAGAACACCTCGACCAACCTGCTGAGCGAGTCCC 2040
 2040 TTTGTTTGGCAACCATTAAGTACTACGTAGAACACCTCGACCAACCTGCTGAGCGAGTCCC 2099
 2041 TGGATGGGGCAATGCCAGTGAAGCATGTTTGGGCAAGAAAGCAACAGTGAAGATCCAAA 2100
 2100 TGGATGGGGCAATGCCAGTGAAGCATGTTTGGGCAAGAAAGCAACAGTGAAGATCCAAA 2159
 2101 TAAGAACTTCAAGCGGATGAAGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
 2160 TAAGAACTTCAAGCGGATGAAGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2219
 2161 TCAAGTAGAGCTGAAAGAGTGAAGGCGCAACTCAAAAGAGCAGAAAGTTTGTTCGGA 2220
 2220 TCAAGTAGAGCTGAAAGAGTGAAGGCGCAACTCAAAAGAGCAGAAAGTTTGTTCGGA 2279
 2221 AGTAACGGAATCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTACGAATATA 2280
 2280 AGTAACGGAATCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTACGAATATA 2339
 2281 TTTGACTCTTCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
 2340 TTTGACTCTTCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2399
 2341 GTTCTTAAAGGAAGTATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2389
 2400 GTTCTTAAAGGAAGTATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 3
 US-10-412-850-9
 ; Sequence 9, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 9
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 US-10-412-850-9

Query Match									
Best Local Similarity 100.0%; Score 2388; DB 16; Length 2451;									
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	TTCTTACGAGTTGGGACCTGTATCAAGCTTAGAACGGTTAAGGAAATAATATCGTGTTCCTTA	60						
Db	60	TTCTTACGAGTTGGGACCTGTATCAAGCTTAGAACGGTTAAGGAAATAATATCGTGTTCCTTA	119						
QY	61	TATAGATGGAAACAAAGCGACGCAAAAAACGGAGAAATTGACTCTCTGATGAGGTTAGCAA	120						
Db	120	TATAGATGGAAACAAAGCGACGCAAAAAACGGAGAAATTGACTCTCTGATGAGGTTAGCAA	179						
QY	121	CGGTGAAGGAATCAATGTCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCA	180						
Db	180	CGGTGAAGGAATCAATGTCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCA	239						
QY	181	TTACATCGCGACCACTATCATTTATTAATGTGTAAAGTTCCTTATGACGCTATGTCA	240						
Db	240	TTACATCGCGACCACTATCATTTATTAATGTGTAAAGTTCCTTATGACGCTATGTCA	299						
QY	241	TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA	300						
Db	300	TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA	359						
QY	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAATACATGTTTACCTTAAAGGATGC	360						
Db	360	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAATACATGTTTACCTTAAAGGATGC	419						
QY	361	TGCCACGGGATACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	420						
Db	420	TGCCACGGGATACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	479						
QY	421	TCAACATCGTGAAGGTGGAACTCCCAAGAAACGATGGTCTGTCGTCGTCGTCGTCGTC	480						
Db	480	TCAACATCGTGAAGGTGGAACTCCCAAGAAACGATGGTCTGTCGTCGTCGTCGTCGTC	539						
QY	481	AGGACGCTATACACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC	540						
Db	540	AGGACGCTATACACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC	599						
QY	541	TGCTGATGCTTATATCGTTCCTCATGAGATCATACCATTAATCTCTAAGAAATGAGTT	600						
Db	600	TGCTGATGCTTATATCGTTCCTCATGAGATCATACCATTAATCTCTAAGAAATGAGTT	659						
QY	601	ATCAGCTAGCGAGTTGGCTCTCGAAGCCCTTCCATCTGGTCGAGGAAATCTGTCAA	660						
Db	660	ATCAGCTAGCGAGTTGGCTCTCGAAGCCCTTCCATCTGGTCGAGGAAATCTGTCAA	719						
QY	661	TTCAAGAACCTATCGCCGACAAATAACGATAACACTTCAAGAACAACTGGGTA	720						
Db	720	TTCAAGAACCTATCGCCGACAAATAACGATAACACTTCAAGAACAACTGGGTA	779						
QY	721	TGTAAGCAATCCAGGAACCTACAAATCTAACACAGCAACACAGCAACCTAACAGTCA	780						
Db	780	TGTAAGCAATCCAGGAACCTACAAATCTAACACAGCAACACAGCAACCTAACAGTCA	839						
QY	781	AGCAAGTCAAAGTAATGACATTCATAGTCTCTTGAACACAGCTCTACAAACCTGCTTG	840						
Db	840	AGCAAGTCAAAGTAATGACATTCATAGTCTCTTGAACACAGCTCTACAAACCTGCTTG	899						
QY	841	TCAACGACATGTAGAAATCTGATGGCCCTTCTTTGATCCAGACAAATCACAGTCGAAC	900						
Db	900	TCAACGACATGTAGAAATCTGATGGCCCTTCTTTGATCCAGACAAATCACAGTCGAAC	959						
QY	901	AGCTAGAGTTTGCAAGTGCACACCGAGATCATACCACTTCTATCCCTTACTCTCAAA	960						
Db	960	AGCTAGAGTTTGCAAGTGCACACCGAGATCATACCACTTCTATCCCTTACTCTCAAA	1019						
QY	961	GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCCTTCGTTATCGTTCAACATG	1020						
Db	1020	GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCCTTCGTTATCGTTCAACATG	1079						

Db 2160 TAAGAACTTTCAAGCGGATGAAGAGCGAGTAGAGAAACACCTCTGTGAGCCAGAGTCCC 2219
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTGTCGAA 2220
Db 2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTGTCGAA 2279
QY 2221 AGTAACGGATTCAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2280
Db 2280 AGTAACGGATTCAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2339
QY 2281 TTTGACTCTTCAATATTATGATACATATAGTATCATCGCAGAGCAGAAAAATTTACTTGC 2340
Db 2340 TTTGACTCTTCAATATTATGATACATATAGTATCATCGCAGAGCAGAAAAATTTACTTGC 2399
QY 2341 GTTGTAAAAAGGAGTAATCCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 2400 GTTGTAAAAAGGAGTAATCCTTCATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 4
US-10-387-783-9
; Sequence 9, Application US/10387783
; Publication No. US2004005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-387-783-9

Query Match 100.0%; Score 2388; DB 16; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTCTACGATTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAAAATAATCGTGTTCCTA 60
Db 60 TTTCTACGATTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAAAATAATCGTGTTCCTA 119
QY 61 TATAGATGAAACAAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCNA 120
Db 120 TATAGATGAAACAAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCNA 179
QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 180
Db 180 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 239
QY 181 TTTACATGGCGACCACTATCAATTTACAAATGGTAAGTTCCTTATGACGCTATCATCAG 240
Db 240 TTTACATGGCGACCACTATCAATTTACAAATGGTAAGTTCCTTATGACGCTATCATCAG 299
QY 241 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAATGA 300
Db 300 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAATGA 359

QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATAACTATGTTTACCTTAAGGATGC 360
Db 360 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATAACTATGTTTACCTTAAGGATGC 419
QY 361 TGCCCAACGGGATAAAGTCCGTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
Db 420 TGCCCAACGGGATAAAGTCCGTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 479
QY 421 TCAACATCTGTAAGGTGGAACCTCAAGAAACGATGGTGTCTGTTGCCCTTGGCACGTTCCGA 480
Db 480 TCAACATCTGTAAGGTGGAACCTCAAGAAACGATGGTGTCTGTTGCCCTTGGCACGTTCCGA 539
QY 481 AGGACGCTATPACTACAGATGATGGTTATATCTTTTAATGCTTCGTATCATAGAGGATAC 540
Db 540 AGGACGCTATPACTACAGATGATGGTTATATCTTTTAATGCTTCGTATCATAGAGGATAC 599
QY 541 TGGTGATGCTTATATCGTTTCTCATGGAGATCATACCATTACATTCCTTAAGATGAGTT 600
Db 600 TGGTGATGCTTATATCGTTTCTCATGGAGATCATACCATTACATTCCTTAAGATGAGTT 659
QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCTATCTTGGTCGAGGAAATCTGTCAAA 660
Db 660 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCTATCTTGGTCGAGGAAATCTGTCAAA 719
QY 661 TTCAAGAAACCTATCGCCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 720 TTCAAGAAACCTATCGCCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 779
QY 721 TGTAAAGCAATCCAGGAACCTAATACTAACAAGCAACAAACAGCAACACTAAGTCA 780
Db 780 TGTAAAGCAATCCAGGAACCTAATACTAACAAGCAACAAACAGCAACACTAAGTCA 839
QY 781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTPACAACTGCCTTTGAG 840
Db 840 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTPACAACTGCCTTTGAG 899
QY 841 TCAACGACATGTGAATCTGATGCGCTTGTCTTGAATCCAGCAGCAAACTCAAGTCAAC 900
Db 900 TCAACGACATGTGAATCTGATGCGCTTGTCTTGAATCCAGCAGCAAACTCAAGTCAAC 959
QY 901 AGCTAGAGGCTTTCAGTGCCACACGAGATCAATACCACTTTCATCCCTTACTCTCAAA 960
Db 960 AGCTAGAGGCTTTCAGTGCCACACGAGATCAATACCACTTTCATCCCTTACTCTCAAA 1019
QY 961 GTCTCAATTTGAAGAACGAATCGCTCGTATTATTCCTTTCCTTTCCTTCAACCATTTG 1020
Db 1020 GTCTCAATTTGAAGAACGAATCGCTCGTATTATTCCTTTCCTTTCCTTCAACCATTTG 1079
QY 1021 GGTACCAATTCAGGCGCAGAACCAAGTCCACACGACTCCGGAACCTAGTCCAGG 1080
Db 1080 GGTACCAATTCAGGCGCAGAACCAAGTCCACACGACTCCGGAACCTAGTCCAGG 1139
QY 1081 CCCGCAACCTCCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 1140
Db 1140 CCCGCAACCTCCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 1199
QY 1141 ACGAAAAAGTTGGGGAAGGATATGTAATCGAAGAAAAAGGCACTCTCTCGTTATGCTTTC 1200
Db 1200 ACGAAAAAGTTGGGGAAGGATATGTAATCGAAGAAAAAGGCACTCTCTCGTTATGCTTTC 1259
QY 1201 GAAAGATTTACCATCTGAAATCTGTTAAATACTTGAAGCAAGTTTATCAAAAACAAGAG 1260
Db 1260 GAAAGATTTACCATCTGAAATCTGTTAAATACTTGAAGCAAGTTTATCAAAAACAAGAG 1319
QY 1261 TGTTCACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGATTTTA 1320
Db 1320 TGTTCACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGATTTTA 1379
QY 1321 TGTAAAGCAATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG 1380
Db 1380 TGTAAAGCAATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG 1439
QY 1381 TAATCTGATTTTCCAGCCCTTAGACAAATATTATAGAACGCTTGAATGATGATCGACTAA 1440


```

Db      1440 TAATCTGATTTCCAGCCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1499
QY      1441 TAAAGAAAAATTTGGTAGATGATTATTGTCATTCTTAGCACCATAATACCATCCAGAGCG 1500
Db      1500 TAAAGAAAAATTTGGTAGATGATTATTGTCATTCTTAGCACCATAATACCATCCAGAGCG 1559
QY      1501 ACTTGGCAACCAAAATTTCAAATTTAGATATCTGAAGACGAAGTTTCGTATTGCTCAATT 1560
Db      1560 ACTTGGCAACCAAAATTTCAAATTTAGATATCTGAAGACGAAGTTTCGTATTGCTCAATT 1619
QY      1561 AGCTGATAAGTATACAACTGATGCTTACATTTTGTGATGATGATATATATCAATGATGA 1620
Db      1620 AGCTGATAAGTATACAACTGATGCTTACATTTTGTGATGATGATATATATCAATGATGA 1679
QY      1621 TGAAGGAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGCTGATTTGGAAGATAG 1680
Db      1680 TGAAGGAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGCTGATTTGGAAGATAG 1739
QY      1681 CCTTCTGATAGGAAAGTTGAGCTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1740
Db      1740 CCTTCTGATAGGAAAGTTGAGCTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1799
QY      1741 TCCATCTCCAGACCCAGATGTTAAAGCAAAATCCAACTGCGATAGTGCAGAGCTATTTTA 1800
Db      1800 TCCATCTCCAGACCCAGATGTTAAAGCAAAATCCAACTGCGATAGTGCAGAGCTATTTTA 1859
QY      1801 CAATCGTGTGAAGGGGAAAAACGAATTTCCACTCGTTCGATCTCCATATATGTTGAGCA 1860
Db      1860 CAATCGTGTGAAGGGGAAAAACGAATTTCCACTCGTTCGATCTCCATATATGTTGAGCA 1919
QY      1861 TACAGTTGAGGTTAAAAACCGTAAATTTGATTAATTCCTATAAGGATCAATACCAATATAT 1920
Db      1920 TACAGTTGAGGTTAAAAACCGTAAATTTGATTAATTCCTATAAGGATCAATACCAATATAT 1979
QY      1921 TAAATTTGCTTGGTTTTCATGATCACATACATACAGCTCCAAATGGCTATACCTTGAAGA 1980
Db      1980 TAAATTTGCTTGGTTTTCATGATCACATACATACAGCTCCAAATGGCTATACCTTGAAGA 2039
QY      1981 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCTCGAGCAAGCTCCCAATCTCTATGA 2040
Db      2040 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCTCGAGCAAGCTCCCAATCTCTATGA 2099
QY      2041 TCGATGGGCAATGCCAGTGAGCATGTGTTAGCAAGAACCCACAGTGAAGATCCAAA 2100
Db      2100 TCGATGGGCAATGCCAGTGAGCATGTGTTAGCAAGAACCCACAGTGAAGATCCAAA 2159
QY      2101 TAAGAACTTCAAAGCCGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Db      2160 TAAGAACTTCAAAGCCGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2219
QY      2161 TCAAGTAGAGACTGAAAAAGTGAAGCCCAACTCAAAGAGCAGAAAGTTTGTCTCGAA 2220
Db      2220 TCAAGTAGAGACTGAAAAAGTGAAGCCCAACTCAAAGAGCAGAAAGTTTGTCTCGAA 2279
QY      2221 AGTAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
Db      2280 AGTAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2339
QY      2281 TTTGACTCTTCAAAATATGATAACAAATAGTATCATGGCAGAGCAGAAAAAATTTACTTGC 2340
Db      2340 TTTGACTCTTCAAAATATGATAACAAATAGTATCATGGCAGAGCAGAAAAAATTTACTTGC 2399
QY      2341 GTTGTAAAGGAAGTAATCTCTCATCTGTAAGTAAGGAAAAAATAAAC 2389
Db      2400 GTTGTAAAGGAAGTAATCTCTCATCTGTAAGTAAGGAAAAAATAAAC 2448

```

RESULT 5

US-10-158-844-94

; Sequence 94, Application US/10158844

; Publication No. US20040029118A1

; GENERAL INFORMATION:

```

;
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
; US-10-158-844-94

```

```

Query Match      100.0%; Score 2388; DB 13; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1   TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTTAAAGAAAAATAATCGTTTCCTA 60
Db      3053 TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTTAAAGAAAAATAATCGTTTCCTA 3112
QY      61  TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
Db      3113 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 3172
QY      121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db      3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232
QY      181 TTCAATGCGCAGCCACTATCATTATTAATCAATGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db      3233 TTCAATGCGCAGCCACTATCATTATTAATCAATGTAAGGTTCCCTTATGACGCTATCATCAG 3292
QY      241 TGAAGAATTTACTCATCAAAAGATCCAAAATATAAGCTTAAAGATGAGGATATTGTTAATGA 300
Db      3293 TGAAGAATTTACTCATGAAGATCCAAAATATAAGCTTAAAGATGAGGATATTGTTAATGA 3352
QY      301 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC 360
Db      3353 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC 3412
QY      361 TGCCACACCGGATACCGTCCCGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 420
Db      3413 TGCCACACCGGATACCGTCCCGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAG 3472
QY      421 TCAACATCGTGAAGGTGGAACCTCCAAGAAAAAGATGGTCTGCTGTCCTTGGCAGCGTTCCGA 480
Db      3473 TCAACATCGTGAAGGTGGAACCTCCAAGAAAAAGATGGTCTGCTGTCCTTGGCAGCGTTCCGA 3532

```


; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-03-884-465A-3

Query Match 57.5%; Score 1374.2; DB 10; Length 2523;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY	1	TTCTTAGAGTGGAGCTGTATCAAGCTAGAGCGTTAAGGAAATATATCGTTTCCTA	60
DB	60	TGCTTATGAATAGGTTTGCATCAAGCTCAAACTGTAAGAAATATATCGTTTCCTA	119
QY	61	TATAGATGGAAACCAAGCGCGCAAAACCGAGAAATTTGACTCTGATGAGTTAGCAA	120
DB	120	TATAGATGGAAACCAAGCGCGCAAAACCGAGAAATTTGACTCTGATGAGTTAGCAA	179
QY	121	CGGTGAAGGAATCAATGCTGAGCAATCGTCTCAAGATACAGACCAAGCTATGTCAC	180
DB	180	CGGTGAAGGAATCAATGCTGAGCAATCGTCTCAAGATACAGACCAAGCTATGTCAC	239
QY	181	TTCAATGGCGACCACTATCATTAATTAATGATGATGATGATGATGATGATGATGATG	240
DB	240	CTCTCATGGAGACCATATCATTAATTAATGATGATGATGATGATGATGATGATGATG	299
QY	241	TGAAGATTAATCAATGAAATCAAACTATAGCTAAAGATGAGGATATGTTAATGA	300
DB	300	TGAAGATTAATCAATGAAATCAAACTATAGCTAAAGATGAGGATATGTTAATGA	359
QY	301	GCTCAAGGGTGGATGATGATCAAGGTAGATGGAATATCAATCGACAAACCAAGAGATAG	420
DB	420	AGCTCATGCGGATAATGTCGCTACAAAGAGAAATCAATCGGCAAAACCAAGAGATAG	479
QY	421	TCAACATCGTGAAGGTGGAATCAAGAAACGATGATGATGATGATGATGATGATGATG	480
DB	480	TAGCATCGTGAAGGTGGAATCAAGAAACGATGATGATGATGATGATGATGATGATG	539
QY	481	AGGACGCTATACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	540
DB	540	GGGACGCTATACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	599
QY	541	TGCTGATGCTTATATCGTTCTCATGAGATCAATACCATTAATTCCTAAGAAATGAGTT	600
DB	600	GGGCGATGCTTATATCGTTCTCATGAGATCAATACCATTAATTCCTAAGAAATGAGTT	659
QY	601	ATCAGCTAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	660
DB	660	ATCAGCTAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	719
QY	661	TTCAAGAACTTATCGCGGCAAAATAGGATACACTTCAAGAAACAACTCGGTAACCTTC	720
DB	720	TTTAAAGAACTTATCGCGGCAAAATAGGATACACTTCAAGAAACAACTCGGTAACCTTC	779
QY	721	TGTAGCAATCCAGGAATCAAAATCACTAACCAAGCAACCAAGCAACCACTAACGATCA	780
DB	780	TGTAGCAATCCAGGAATCAAAATCACTAACCAAGCAACCAAGCAACCACTAACGATCA	839
QY	781	AGCAAGTCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	840
DB	840	AGCAAGTCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	899
QY	841	TCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	900
DB	900	TCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	959
QY	901	AGCTAGAGGTGTTGAGTGCCACAGGAGATCAATACCATTTCTATCTCTCAAT	960

DB	960	CGCCAGAGGTGTAGCTGCTCCCTCATGTGTAACTTACCACTTTATCCCTTATGAACAAAT	1019
QY	961	GTCTGAATTTGAAGAAACGAATCGCTCGTATTATTATTTCCCTTCTGTTATCGTTCAACCAATG	1020
DB	1020	GTCTGAATTTGAAGAAACGAATCGCTCGTATTATTATTTCCCTTCTGTTATCGTTCAACCAATG	1079
QY	1021	GGTACCAGATTTCAAGCGCGCAAAACCAAGTCCACCAACCGACTCCGGAACCTAGTCCAGG	1080
DB	1080	GGTACCAGATTTCAAGCGCGCAAAACCAAGTCCACCAACCGACTCCGGAACCTAGTCCAGG	1139
QY	1081	CCCGCAACCTGCGACCAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT	1128
DB	1140	TCCGCAACCTGCGACCAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT	1199
QY	1129	TAGTCAGCTGTGTACGAAAGTTGGGGAAGGATATGTTTCCGAAAGCAATCCAATTTGATGAAATTTGGT	1188
DB	1200	CAAGAAGCTGTTTCCGAAAGTTAGCGGATGTTTGTCTTTGAGGAGATGAGGATTTCTCG	1259
QY	1189	TTATGCTTTTGGCGAAGATTTTACCATCTGAACTGTTTAAATAATCTTCAAAAGCAAGTTTATC	1248
DB	1260	TTATATCCGAGCCCAAGATCTTTTCCGAGAAACAGCAGCAGGCAATGATAGCAACTGGC	1319
QY	1249	AAACAGAGAGGTGTTTTCACACACTTTTAACTGCTTAAATAATCTTCAAAAGCAAGTTTATC	1308
DB	1320	CAAGCAGGAAAGTTTATCTCATAGCTTAGGAGCTTAAAGAACTGACCTCCATCTAGTGA	1379
QY	1309	CCAAGATTTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGN	1368
DB	1380	TCGAGATTTTCAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTTGA	1439
QY	1369	AAATAAGGGTCTGTAATTTCTGATTTCCAAGCTTTAGACAAATTTATTAAGACGCTTGAATGA	1428
DB	1440	TAATAAAGGTCTGCAAGTTGATTTTGAAGCTTTGGATAACCTGTTGAAACGACTCAAGGA	1499
QY	1429	TGAATCGACTTAATAAAGAAATTTGTTAGATGATTTATTTGGCATTTCTAGCAACCAATTC	1488
DB	1500	TGCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTTCTTGCCTTTCTAGCTCCGATTCG	1559
QY	1489	CCATCCAGAGCGACTTTGGCAAAACCAAAATTTCTCAAAATTTAGTATACCTGAAGACCAAGTTTCG	1548
DB	1560	TCATCCAGAGCTTTAGGAAACCAAAATGGCAAAATTCCTACACTGATGATGAGATTTCA	1619
QY	1549	TATGCTCAATTAGCTGATAAGTATACAAGTATACAAGTATGATGTTTATTTTATGATGAAATGA	1608
DB	1620	AGTAGCAAGTTGGCAGGCAAGTACACACAGAAAGCGTTATATCTTTGATCTCTGTTGA	1679
QY	1609	TATAATCAGTGTAGAGGATGTCATATGTAACGCTCATATGCGGCTATATGCGGCTAGTCACTGGAT	1668
DB	1680	TATAACCAAGTGTAGGAGGATGCTTATGTAATCTCCACATATGACCCATAGCCACTGGAT	1739
QY	1669	TGGAAAAGATAGCTTTCTGATAAGGAAAGTTGCGAGCTCAAGCTCATACTAATAAGAA	1728
DB	1740	TAAAAAAGATGTTTGTCTGAAGCTTGAGAGCGGCGAGCCCAAGCTTATGCTTAAAGAGAA	1799
QY	1729	AGGTATCTTATCTCCATCTCCAGACGAGATTTAAAGCAAAATCCAATCTGGAGATAGTGC	1788
DB	1800	AGGTTTGACCCCTCTCTCCAGACGATCCATCAGGATTTAGGAAATCTAGGCGCAAGGAGC	1859
QY	1789	AGCAGCTATTTACATCGTGTGAAAGGGGAAACGAATTCACCTGCTTCCATTA	1848
DB	1860	AGAGCTATCTACAAACCGGCTGAAAGAGAGCTTAAAGAGGTGACCTTGTATGCTTGA	1919
QY	1849	TATGTTGAGCATACAGTTGAGTTTAAAGCGTAAATTTTATTTTATTTCTCTAAGATCA	1908
DB	1920	CAATCTTCAATATCTGTAGAGTCAAAACGTTAGTTTATTAATCACTTATTAAGCA	1979
QY	1909	TTACATTAATTTAAATTTGCTTTGATGATCACACATACAAAGCTTCCAAATGGCTA	1968
DB	1980	TTACCATTAATCAATTTGAGTGTGAGTGTGAGGAGGCTTTTATGAGGACCTTAAAGGGA	2039
QY	1969	TACCTTGAAGATTTGTTGCGAGATTAAGTATCTAGTAAACACCTTACGACCACTGCC	2028

Db 2040 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAACGTC 2099
QY 2029 ACATTTCTAATGATGGATGGGGCAATGCGCAGTGAAGCATGTGTTAGGCGAAGAACCCACAG 2088
Db 2100 GCATTCAGATAATAGTGTGTTGTTAAACGCTAGCGACCATGTTCAAGAGAAACAAAAATGGTCA 2159
QY 2089 TGAAGATCCAAATAGAACTTCAAGCGGATGAAGAG----- 2125
Db 2160 AGCTGATACCAATCAAAACGGAAACCAAGCGAGGAGAAACCTCAGACAGAAACCTGA 2219
QY 2126 ----- 2125
Db 2220 GGAAGAAACCCCTCGAGNAGAGAACCCACAAAGCGAGAACACAGAGTCTCCAAACCAAC 2279
QY 2126 -----CCAGTAGAGGAAACACCTGCTGAGCCGAGAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2280 AGAGGAACCAAGAGAAATCACCAGAGGAATCAGAGAACTCAGAGTCGAGACTGAAAA 2339
QY 2179 AGTAGAGCCCAACTCAAGAGAGCAGAAAGTGTTCCTGCGAAAGTAAACCGGATTCAGTCT 2238
Db 2340 GGTTGAAGAAACCTGAGAGAGGTGAAGATTACTTGGAAATCCAGGATCCAAATAT 2399
QY 2239 GAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATTAATTGACTCTTCAAAATAT 2298
Db 2400 CAAGTCCCAATGCCAAGAGACTCTCACAGGATTAAGAAATTAATTACTATTTGGCACCCA 2459
QY 2299 GGATAACAATAGTATCATGGCAGAGCAGAGAAATTAATTGCTGGTTTAAAGGAGTAA 2358
Db 2460 GGACAACAATACTATTATGGCAGAGCTGAAGAACTATTGGCTTTTAAAGGAGAGTAA 2519

RESULT 7

US-09-884-465A-4
; Sequence 4, Application US/09884465A
; Publication No. US20030077293A1

GENERAL INFORMATION:

; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2647
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae

US-09-884-465A-4

Query Match 57.5%; Score 1374.2; DB 10; Length 2647;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60
Db 104 TGGTTATGACTAGGTTTGCATCAAGCTCAAGCTGTAAAGAAAAATAATCGTGTTCCTA 163
QY 61 TATAGATGGAACAAACGAGCGCAACAAACGAGAAATTTGACTCCCTGATGAGTTAGCAA 120
Db 164 TATAGATGGAACAAACGAGCGCAACAAACGAGAAATTTGACTCCCTGATGAGTTAGCAA 223
QY 121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAGATAAACAGCAAGGCTATGTCAC 180
Db 224 GCGTGAAGGAATCAACGCCCAACAAATCGTCATCAGATTAACGATCAAGTTATGTGAC 283
QY 181 TTCACATGGCGCACTATCATTTACAAATGAGGTTTCTTATGACGCTATCATCAG 240

Db 284 CTCTCATGGAGACCATTTATCAITTAATAATGGCAAGTCCCTTTATGATGCCATCATCAG 343
QY 241 TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATGTTTAATGA 300
Db 344 TGAAGAGCTCTCTCATGAAGATCGGAATATCAGTTGAAGGATTCAGACATTTGTCATGA 403
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 404 AATCAAGGGTGGTATGTCATTAAGGTAAACGGGTAAATACTATGTTTACCTTAAGGATGC 463
QY 361 TGCCACGCGGATTAACGTCCTACAAAGAGGAATCAATCGACAAAAAACAAGACATAG 420
Db 464 AGCTCATGCGGATTAATGTCCTACAAAGAGGAATCAATCGGCAAAAAACAAGACATAG 523
QY 421 TCAACATCGTGAAGGTGGAACTCCAAAGAACGATGCTGTGCTTGGCTTGGCACGCTTCGCA 480
Db 524 TCAGCATCGTGAAGAGGGGACTTCAGCAACGATGGTGGGTAGCCCTTGCACGTTCAACA 583
QY 481 AGGACGCTATACATGATGATGTTTAAATGCTTCTGATATCATAGAGGATAC 540
Db 584 GGGACGCTACACACAGATGATGTTTATCTTCAATGCAATCTGATATCATCGAGATAC 643
QY 541 TGGTGAATGCTTATATGTTCTCATGGAGATCATTAACCATTAATCTCTTAAGAAATGAGTT 600
Db 644 GGGGATGCTCTATATGTTCTCATGGAGATCATTAACCATTAATCTCTTAAGAAATGAGTT 703
QY 601 ATCAGTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGCTGAGGAAATCTGTCAAA 660
Db 704 ATCAGTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGCTGAGGAAATCTGTCAAA 763
QY 661 TTCAAGAACCTTATCGCGACAAAAATAGCGATTAACATTCAGAGAACAACTGGTACCTTC 720
Db 764 TTTAAGAACCTTATCGCGACAAAAATAGCGATTAACATTCAGAGAACAACTGGTACCTTC 823
QY 721 TGTAAAGCAATCCAGGAACTACAAATACTAAACAGCAACAAACACAACTAAACAGTCA 780
Db 824 TGTAAAGCAATCCAGGAACTACAAATACTAAACAGCAACAAACACAACTAAACAGTCA 883
QY 781 AGCAAGTCAAAAGTAAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG 840
Db 884 AGCAAGTCAAAAGTAAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG 943
QY 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 944 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1003
QY 901 AGCTAGAGGTGTTGAGTGGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 1004 CGCCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063
QY 961 GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTGGTTATGCTGTTTCAACCAT 1020
Db 1064 GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTGGTTATGCTGTTTCAACCAT 1123
QY 1021 GGTACAGATTTCAAGCCAGAACCAACAGTCCCAACCGACTCCCGAACCTAGTCCAG 1080
Db 1124 GGTACAGATTTCAAGCCAGAACCAACAGTCCCAACCGACTCCCGAACCTAGTCCAG 1183
QY 1081 CCGGCAACCTGCACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGTT 1128
Db 1184 TCGGCAACCTGCACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGTT 1243
QY 1129 TAGTCAGCTGGTACGAAAAAGTTGGGAAGGATGATGATTTGGAAGAAAAAGGCAATCTCTCG 1188
Db 1244 CAAAGAGCTGTTTCGAAAAAGTAGGCGATGTTATGCTTTTGGAGAGATGGAGTTTCTCG 1303
QY 1189 TTATGCTTTTGGAAAGATTTACCATCTGAACCTGTTTAAATCTTTGAAAGCAAGTTATC 1248
Db 1304 TTATATCCCGCAAGATCTTTTCAAGAGAAACAGCAGGCAATGATGAGCAAACTGGC 1363
QY 1249 AAAACAGAGAGTGTTCACACACTTTTAACTGCTAAAAAGAAAAATGTTCTCTCTGTTGA 1308

Db 1364 CAAGCAGGAAGTTTATCTCATAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAGTGA 1423
Qy 1309 CCAAGAAATTTTATGATAAGCATATAATCTGTGTTAACTGAGGCTCATAAAGCCTTGTTGN 1368
Db 1424 TCGAGAAATTTTACATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTTGA 1483
Qy 1369 AAATAAGGTCGTAAATCTGATTTCCAAAGCCCTTACACAAATTTATAGAACGCTTGAATGA 1428
Db 1484 TAATAAAGGTCGCAAGTTGATTTGAGGCTTTGGATAACCTGTTGGAAACGACTCAAGGA 1543
Qy 1429 TGAATCGACTAATAAGAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACCATAATPAC 1488
Db 1544 TGTCTCAAGTATGAAGTCAAGTTAGTGGATGATTTCTTGCCTTCTTAGCTCCGATTCG 1603
Qy 1489 CCATCCAGAGCGCTTGGCAAAACCAATTTCTCAAAATTTAGTATCTGAAGCAAGTTTCG 1548
Db 1604 TCATCCAGAAGTTTAGAAACCAATGCGCAATTTACCTACACTGATGATGATTTCA 1663
Qy 1549 TATTGCTCAATAGCTGATTAAGTATACAAAGTTCAGATGGTTTACATTTTGTATGACATGA 1608
Db 1664 AGTAGCCAAAGTTGGCAGGCAAGTACACACAGAGAGCGGTTATATCTTTGTATCCTCGTA 1723
Qy 1609 TATAATCAGTATGAAGGATGATATGTAAAGCCCTCATATGGGCGCATAGTCACTCGAT 1668
Db 1724 TATAACCGATGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1783
Qy 1669 TGGAAAGATAGAGCTTTCTGATAAGGAAAAAGTTGACAGTCAAGCCCTATATCTAAAGAAA 1728
Db 1784 TAAAAAAGATGTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAAGAGAA 1843
Qy 1729 AGGTATCTACCTCCATCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGC 1788
Db 1844 AGGTTTGAACCCCTCTTCGACAGACCATCAGGATTCAGGAATTAATCAGGCAAAAGGAGC 1903
Qy 1789 AGCAGCTATTACAAATCGTGTGAAGGGGAAACAGATTCCTGCTGCTTCCATTA 1848
Db 1904 AGAAGCTATCTACACCGGCTGAAGAGCAGTAAAGAGGTGCCACTTGTATGCTATGCTTA 1963
Qy 1849 TATGTTTGAGCATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGATCA 1908
Db 1964 CAATCTTCAATATCTGTAGAGTCAAAACGGTAGTTTAAATCATACCTCATTTATGACCA 2023
Qy 1909 TTACCTAATATTAATTTGCTTGTGATGATCACATACAAAGCTCCAAATGGCTA 1968
Db 2024 TTACCAATAACATCAATTTGAGTGGTTTGACGAAGCCCTTTATGAGSCACCTTAAGGGTA 2083
Qy 1969 TACCTTGAAGATTTGTTGCGAGATTAAGTACTAGTAGAAGACCTTGACGAGCTCC 2028
Db 2084 TACTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACTCC 2143
Qy 2029 ACATTTCTAATGATGGGCGCAATGCCAGTGAAGCATGTTGAGGCAAGAAAGACACAG 2088
Db 2144 GCATTCAGATAATGTTTGTGTTAAAGCTAGGACCATGTTCAAGAAACAAATAATGGTCA 2203
Qy 2089 TGAAGTCCAAATAGAACTTCAAAGCGGATGAAGAG----- 2125
Db 2204 AGCTGATACCAATCAAAACGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTCA 2263
Qy 2126 ----- 2125
Db 2264 GGAAGAAACCCCTCGAGAAGAGAAACCAAAAGCGAGAAACAGAGTCTTCCAAAACCAAC 2323
Qy 2126 -----CCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCCTCAAGTAGAGACTGMAAA 2178
Db 2324 AGAGGAACCAAGAAGAAATCACAGAGGAATTCAGAAGAACCTCAGGTCGAGACTGMAAA 2383
Qy 2179 AGTAGAGCCCACTCAAGAAGCAGAAAGTTTGTCTTGGAAAGTAAACGATTTCTAGTCT 2238
Db 2384 GGTGGAAGAAAACTCGAGAGAGGCTGAAGATTTTCTTGGAAAAATCCAGGATCCAAATAT 2443
Qy 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATATTTGATCTTTCAAATAT 2298
Db 2444 CAAGTCCCAATGCAAGAGACTCTCACAGGATTTAAAAAATAATTTACTATTGTCACCCA 2503

Qy 2299 GGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTGCTTGTAAAAAGGAGTAA 2358
Db 2504 GGACAACAATACTATTATGGCAGAGCTGAAAAACTATTGCTTTATTAAAGAGAGTAA 2563

RESULT 8

US-09-769-787-206
; Sequence 206, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; FILE OF INVENTION: Proteins
; TITLE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-206

Query Match 42.0%; Score 1003.8; DB 10; Length 2481;
Best Local Similarity 66.0%; Pred. No. 6.5e-233;
Matches 1623; Conservative 0; Mismatches 693; Indels 144; Gaps 6;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAGGAAA---TAATCGTGTTC 57
Db 60 TTCTTATGAGCTTGGTGTGTCACCAAGCTGTCAGGATAAGAAAGTCTAATCGAGTTGC 119
Qy 58 CTATATAGATGGAACAAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATCAGGTTAG 117
Db 120 TTATATAGATGGTGTATCAGGCTGGTCAAAAGGAGAAACTTGACACAGATGAAGTCAG 179
Qy 118 CAGCGTGAAGAAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGT 177
Db 180 TAAAGGGAGGGGATCAACCGCGAAATAATCGTCATCAAGATTACGGATCAAGGTTATGT 239
Qy 178 CACTTCACATGCGACCACTATCATTTATACATGTTAGGTTTCTTATGAGCTATCAT 237
Db 240 GACTCTCATGAGAGACCATTAATTTATCTATTAATGCAAGGTCCTTATGATGCCATCAT 299
Qy 238 CAGTGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAAATTAATCAGTTGAAGGATTCAGACATTCNA 359
Qy 298 TGAGCTCAAGGTTGATATGTTATCAAGTAGATGGAATACTATGTTTACCTTTAAGGA 357
Db 360 TGAATCAAGGTTGTTTATGTCATCAAGTAGACGCAAAATACTATGTTTACCTTTAAGGA 419
Qy 358 TGCTCCCAAGCGGATAACGTCGTACAAAGAGGAAATCAATCCGAAAAACAAGACA 417
Db 420 TGCAGCTCATGCGGATAATATTGCGACAAAAGAGAGATTAAACGTCAGAAGCAGGAACG 479
Qy 418 TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGCTGTTGCTTGGCAGCTTC 477
Db 480 CAGTCATAATCACGGGTGAGGAGCT-----ACGATCATGCTAGTAGCTGAGCCAGAGC 533
Qy 478 GCAAGGAGCTTATCTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGA 537
Db 534 CCAAGGAGCTTATCAACGGATGATGGGTATCTTCAATGATCTGATATCATTTAGGA 593
Qy 538 TACTGGTGATCTTATATCGTTTCTCATGGAGATCATACCATTCATTCCTCAAGATGA 597
Db 594 CACGGTGATGCTTATATCTGTTCTCTCAGCGGCGCAATTAACCATTTACATTCCTCAAGATGA 653

598 GTTATCAGCTAGAGTGGTCTGCTGAGAGGCTTCCTATCTGTCGAGGAATCTGTC 657
 1666 GATTGGAAGAAGTATGAGCTTTCTGATAGGAAGTTCGAGCTCAAGCTTATCTAAGA 1725
 1698 GATTAAAAAGATAGTGTCTGAGAGCTGAGAGCGGCGAGCCGCTTATCTAAGA 1757
 1726 AAAGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCCAACTGAGATAG 1785
 1758 GAAAGGTTGACCCCTCTCTCGACAGCAATCAGGATTCAGGAATACTGAGGCAAAAG 1817
 1786 TGACAGAGCTATTTACAACTCGTGTGAAAGGGGAAAAACGAAATTCCTCGTTCAGCTTCC 1845
 1818 AGCAGAGCTATCTCAACCGCTGAAAGCAGCTAAGAGGTCGCACTTGTATCGTATGCC 1877
 1846 ATATATGTTGAGCATACAGTTCAGGTAAAAAGGTAATTTGATTTATCTCTATAAGGA 1905
 1878 TTCAATCTTCAATATATCTAGAGTCAAAACGGTAGTTTAAATCATCTCTATTATGA 1937
 1906 TCATTACCAATAATTAATTTGCTGTTTGTATGATCACACATACAAAGCTCCAAATGG 1965
 1938 CCATTACCAATAATTAATTTGCTGTTTGTATGATCACACATACAAAGCTCCAAATGG 1997
 1966 CTATACCTTGGAAAGTATTTGTTGCGAGCATTAAGTACTAGTGAACACCTCTGACGACG 2025
 1998 GTATATCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTTCGAACATCCAAACGACG 2057
 2026 TCACATCTTAATGATGATGGGCAATGCGAGTGAACATGTTAGGCAAGAAAGACCA 2085
 2058 TCCGATTCAGATAATGTTTGGTAACTGCTAGCGCCATGTTTCAAGAAACAAATAATGG 2117
 2086 CAGTGAAGATCCAAATTAAGACTTCAAGGGGATGA----- 2121
 2118 TCAAGCTGATACCAATCAACCGAANAACNAGCGAGGAGAACTCAGACAGAAAAACC 2177
 2122 ----- 2121
 2178 TGAGGAAGAACCCCTCGAGAGAGAAACCGAAAGCGAGAAACCGAGTCTCCAAAAACC 2237
 2222 ---AGAGCAGTAGAGGAACACCTGTGAGCCGAGAGTCCCTCAAGTAGAGACTGAAAA 2178
 2238 AACAGAGAACCCAGAGAAATCACCAGAGAAATCAGAGAACTCAGAGTCTGAGACTGAAAA 2297
 2179 AGTAGAGCCCAACTCAAGAGAGAGAGTGTTCGTCGGAAGTAAACGAGTCTTAGTCT 2238
 2298 GGTGAGAAAGAACTGAGAGAGCTGAGATTTACTTGGAAAAATCCAGGATCCAAATAT 2357
 2239 GAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATTAATTTGACTCTTCAAAATAT 2298
 2358 CAAGTCCATGCGAAGAGACTCTCAGAGATTTAAAAAATAATTTACTATTITGGCAACCA 2417
 2299 GATTAACAAATAGTATCATGCGAGAGAGCAAAAAATTTCTGCTGTTTAAAGAGAGTAA 2358
 2418 GGACAACAATATCTATTATGGCAGAGCTGAAAAACTATTGGCTTTATTAAAGAGAGTAA 2477

RESULT 9

US-10-412-862-11
 ; Sequence 11, Application US/10412862
 ; Publication No. US20040052781A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-685
 ; CURRENT APPLICATION NUMBER: US/10/412,862
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 11

; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-862-11
 Query Match 41.5%; Score 991; DB 13; Length 2531;
 Best Local Similarity 65.7%; Pred. No. 8.5e-230;
 Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;
 1 TTCTTACAGTTGGGACGTGATCAGCTAGACGTTAAGGAAA---TAATCGHGTTC 57
 60 TTCCTATGAGCTTGACGCTTACAGCTTGGTCAAGGATGAAGAAAGAGTCTAATCCGAGTTGC 119
 58 CTATATAGATGGAAGAAACGAGCGACGCAAAACGAGAAATTTGACTCTCTGATGAGGTTAG 117
 120 TTATATAGATGCTGATCAGGCTGGTCAAGGCAAGAACTTGACACCAAGATGAAGTCAG 179
 118 CAAGCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
 180 TAAGAGGAGGGGATCAACGCGCAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT 239
 178 CACTTTCATGCGGACCACTATCATTTATCAATGTTAAGGTTCTTATGACGCTATCAT 237
 240 GACTCTCATGAGACCAATTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 299
 238 CAGTGAAGAATTACTCATGAAGAATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAA 297
 300 CAGTGAAGAGCTCTCATGAAGATCCGAATTTATCAGTTGAGGATTCAGACATTTGCAA 359
 298 TGAGGTCAGAGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGA 357
 360 TGAATCAAGGGTGGTATGTCATTAAGGTAAACGGTAAATCTATGTTTACCTTAAGGA 419
 358 TGCTGCCACCGGATAAAGTCCGTCACAAAGAGGAATCAATCGCAAAACCAAGAGCA 417
 420 TSCRGCTCATCGGATATTAATTCGCAAAAGAGAGATTAAGACGTCAGAGGAGGAACG 479
 418 TAGTCAACATCGTGAAGGTGAACTCCAAAGAAACGATGGTGTCTGCTTGGCCCTTGCACGCTTC 477
 480 CAGTCAATATCAT-----AACTCAAGAGCAGATTAATGCTGTCTGCACGCCAGAGC 530
 478 GCNAGAGCGTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
 531 CCAAGAGCGTTATACAAAGGATGATGGTATATCTTCAATGCATCTGATATCATTCAGGA 590
 538 TACTGTGATGCTTATATCTCTCATGAGATCATTAACCATTAATCTTCTTAAGATGA 597
 591 CAGGCGTATGCTTATCTGTTCTCAGCGGACCAATTAACCATTAATCTTCTTAAGATGA 650
 598 GTTATCAGCTAGCGATGCTGCTGCAAGGCTTCTCTATCTGTCGAGGAATCTGTC 657
 651 GTTATCAGCTAGCGATGCTGCTGCAAGGCTTCTGGAATGG----- 695
 658 AAATTCAGAACTTATCGCGCAAAATAGGATTAACCTTCAAGAACAACTGGGTACC 717
 696 -GAAGCAGGATCTGCTCTCTTCTTCAAGTCTAGTTATTAATGCAATCCAGCTCAACCAA 754
 718 TTCTGTAAAGCAATCCAGGAATCAAAATCTAACCAAGCAACACAGCAACACTTAACAG 777
 755 GATTGTGAGAACCAATCT-----GACTGTCACTTCAACTTA 794
 778 TCAAGCAAGTCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
 795 TCATCAAAATCAAGGGGAAAACATTTCAAGGCTTTTACGTGAATGTTATGCTTAACCCCTT 854
 838 GAGTCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
 855 ATCAGAACCCATGTTGAATCTGATGGCTTTATTTTCGACCCAGCGCAATCAAGTCG 914
 898 AACAGCTAGAGGTGTTGAGTCCACAGGAGATCATTAACCACTTATCCCTTACTCTCA 957
 915 AACCCCGAGAGGTGATGCTGCTCCATGTTGATTAACCACTTATCCCTTATGAACA 974

QY 958 AATGTCTGAATTTGGAAGAACGAATCGCTGATTTATTCCTCTTCTGTTATCGTTCAAAACA 1017
 DB |||||
 975 AATGTCTGAATTTGGAAGAACGAATTTGCTGATTTATTCCTCTTCTGTTATCGTTCAAAACA 1034
 QY 1018 TTGGGTACAGATTTCAAGGCCAGACCAACCAAGTCCCAACAGCTCCGGAACTTAGTCC 1077
 DB |||||
 1035 TTGGGTACAGATTTCAAGGCCAGACCAACCAAGTCCCAACAGCTCCGGAACTTAGTCC 1094
 QY 1078 AGGCCCGCAACTGTCACCAAAATCTTAAATAGACTCAAAATCTTCTTCTGTTATCGTCC 1137
 DB |||||
 1095 AAGTCCGCAACAGCTCCAGCAATCCAAATTTGATGGGAA-----ATTGGTCAAAAGAAC 1148
 QY 1138 GGTACGAAAAGTTGGGGAAGGATATGTTATCGAAAGAAAGGCGCATCTCTGTTATGTTCT 1197
 DB |||||
 1149 TGTTTCGAAAGTAGGCGATGTTATGTTTGGAGAGAAATGGAGTTTCTCGTTATATCC 1208
 QY 1198 TGGCAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTTATCAAAACAAGA 1257
 DB |||||
 1209 AGCCAGAGATCTTTTCAGCAGAAACAGCAGCAGGAGCAATTTGATAGCAAACTGGCACAAGCA 1268
 QY 1258 GAGTGTTCACACACTTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCGACCAAGATTT 1317
 DB |||||
 1269 AAGTTTATCTCATAGCTTAGCACTAGAAAACTGACCTCCATCTAGTATCGAGAAATTT 1328
 QY 1318 TTATGATAAGCAATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGAAATAAAGG 1377
 DB |||||
 1329 TTAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGTATATAAGG 1388
 QY 1378 TCGTAATCTGATTTTCAAGCCCTTAGCAAAATTTATTAGAACGCTTGAATGATGATCGAC 1437
 DB |||||
 1389 TCGCAAGTTGATTTTGGGCTTTGGGATAACCTGTTGGAACGACTCAAGGATGTTCTCAAG 1448
 QY 1438 TAATAAGAAAAATTTGTTAGATGATTTATTGCGCAATCTTAGCACCCTTATGTTATGTTCTCA 1557
 DB |||||
 1449 TGATAAGTCAAGTTAGTGGAGATATTCTTGCTTCTTAGCTCCGATTCGTTATCGTCAATCCAGA 1508
 QY 1498 GCACCTTGGCAAAACCAAAATTTCTCAATTTAGTATATCTGAAAGACGAGTTCGTTATGTTCTCA 1557
 DB |||||
 1509 ACCTTTAGGAAAAACCAAAATTCGCAAAATTAACCTACACTGATGATGAGATTCAGTACGCA 1568
 QY 1558 ATTAGCTGATTAAGTATACACGCTCAGATGGTTACATTTTGTGATGAACATGATATAATCG 1617
 DB |||||
 1569 GTTGGCAGCAAGTACACAGCAGACGCGTTATATCTTTGATCCTCGTGATATAACAG 1628
 QY 1618 TGATGAAGGAGATGCTATATGTAACGCTCATATGGGCCATAGTCACTGATTCGAAAGA 1677
 DB |||||
 1629 TGATGAGGGGATGCTATGTAACCTCAATATGACCCATAGCCATGAGCTTGGATTTAAAAAGA 1688
 QY 1678 TAGCCTTTCTGATAGGAAAAAGTTGACGCTCAAGCTTATCTTAAAGAAAAAGGTTATCCCT 1737
 DB |||||
 1689 TAGTTGTCTGAAGCTTGAGAGCGCAGCCCGGCTTATGCTTAAAGAGAAAGGTTTGAC 1748
 QY 1738 ACTTCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGGCAGAGCTAT 1797
 DB |||||
 1749 CCTTCTTCCAGACCACTCAGATTCAGGAAATTAAGGCAAAAGGAGAGAGAGAGCTAT 1808
 QY 1798 TTCAACTCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCCACTTCCATATATGTTTCA 1857
 DB |||||
 1809 CTACAACCGMTGAAAGCAGCTAAGAGGTGCACTTGATCTGATGCTTATGCTTATCAATCTTCA 1868
 QY 1858 GCATCAGTTGAGGTTTAAACCGGTAATTTGATTAATTCCTCAATAAGGATCAATTAACATAA 1917
 DB |||||
 1869 ATATATCTGTAAGGTCAAAAAACGGTAGTTTAAATCATACCTCATTTATGACCAATACCAAT 1928
 QY 1918 TATTAAATTTGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
 DB |||||
 1929 CATCAATTTGAGTGGTTTGGAGAGGCTTTATGAGGCACTTAAGGCGATATACCTTTGA 1988
 QY 1978 AGATTGTTGCGAGGATTAAGTACTAGTGAACACCCCTGACCAACCTGCAATCTTCA 2037
 DB |||||
 1989 GGATCTTTGGGAGCTGTCAGAGTACTATGTCGAACATCCAAACGAACTGCGCATTTCA 2048
 QY 2038 TGATGGATGGGGCAATGCCAGTGAAGATGTTGTTAGGCAAGAAAGAACACAGTGAAGATCC 2097

Db 2049 TAATGGTTTGGTAAACGCTAGCAGCACCATTGTTCAAGAAACCAAAATGGTCAAGCTGATAC 2108
Qy 2098 AAATAAGAACTTCAAAAGCGGATGA----- 2121
Db 2109 CAATCAACCGAAACCAACGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAAC 2168
Qy 2122 -----AGAGCCAGT 2130
Db 2169 CCCTCGAAGAGAAACCGCAAGCGAGAAACCCAGAGTCTCCAAAACCAACAGAGGAACC 2228
Qy 2131 AGAGGAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAAGCTAGAAGCCCA 2190
Db 2229 AGAAGAAATCAACGAGGAATCAGAAAGAACCTCAGTGCAGACTGAAAGGTTGAAGAAA 2288
Qy 2191 ACTCAAGAAAGACGAAAGTTTGGTTGCGAAAGTAACCGATTCAGTCTGAAAGCCAAATGC 2250
Db 2289 ACTGAGAGAGCTGAAGATTACCTTGGAAAAATCCAGGATCCAAATTAATCAAGTCCAATGC 2348
Qy 2251 AACGAAACTCTAGCTGGTTTACGAAATATTTGACTCTTCAAAATTTAGGATAACAATAG 2310
Db 2349 CAAGAGACTCTCAGAGATTAAAGAAATATTTACTATTTGGCACCCAGGACCAACAATAC 2408
Qy 2311 TATCATGCGAAGCAGAAAAATTTACTTGCCTTCTGTTAAAAAGGAAGTAA 2358
Db 2409 TATTATGCGAAGCTGAAAAACTATTGGCTTTTAAAGGAGAGTAA 2456

RESULT 10
US-10-412-850-11
; Sequence 11, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie E.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-11

Query Match 41.5%; Score 991; DB 16; Length 2531;
Best Local Similarity 65.7%; Pred. No. 8.5e-230;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;
Qy 1 TTCTTACGATGGGACTGTATCAAGCTAGAACGGTTAAGGAAA-----TAATCGTGTTC 57
Db 60 TTCTTACGATGGGACTGTATCAAGCTAGAACGGTTAAGGAAA-----TAATCGTGTTC 119
Qy 58 CTATATAGATGGAACCAACAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAG 117
Db 120 TTATATAGATGATGATCAGCTGGTCAAAAGGAGAAACTTGACACCAAGATGAAGTCA 179
Qy 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGCAAGGCTATGT 177
Db 180 THAAGGGAGGGGATCAACGCCGAAACAAATTTGTTATCAAGATTAACGATCAAGGTTATGT 239
Qy 178 CACTTCACATGGCGACCACTATCATTTTACATGTAAGTTCCTTTATGACCTATCAT 237
Db 240 GACCTCTCTGAGAGACCATATATCTATTAATGCAAGGTTCCCTTATGATGCCATCAT 239

Qy 238 CAGTGAAGAAATTACTCATGAAGATCCAAACTATTAAGCTAAAGATGAGATATTCGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGATTCAGACATTCGTCAA 359
Qy 298 TGAGGTCAAGGGTGGATATCTTATCAAGGTAGATGGAATACTATGTGTTTACCTTAAGGA 357
Db 360 TGAATCAAGGGTGGTATGTCATTAAAGTTAAACGGTAAATATACTATGTGTTTACCTTAAGGA 419
Qy 358 TGCTGCCACGCGGATAAAGTCCTGTACAAAAGAGGAAATCAATCGACAAAAACAAGGCA 417
Db 420 TGRGCTCATGCGGATAATATTTCGGACAAAAGAGAGATTAACCGTCAGAAAGGAGAACG 479
Qy 418 TACTCAACATCTGTGAAGGTGGAACCTCAAGAAAAGTGTGCTGTTGCGCTTGCACGCTC 477
Db 480 CAGTCATAATCAT-----AACTCAAGAGCAGATAATGCTGTGCTCGAGCCAGC 530
Qy 478 GCAGGAGCGCTATACACAGATGATGTTATATCTTTAAATGCTTCTGATATCATAGAGGA 537
Db 531 CCAAGGAGCGTTATACAAAGGATGATGGGTATATCTTCAATGTCATCTGATATCATTTAGGA 590
Qy 538 TACTGTGTGATGCTTATATCGTTCCTCATGGAGATCATTAACATTAATTCCTTAAGAAATGA 597
Db 591 CACGGGTGATGCTTATATCGTTCCTCAGCGGACCATTAACCATTAACATTCCTTAAGAAATGA 650
Qy 598 GTTATCAGCTAGCAGTGTGCTGCTGAGAGCGCTTCTATCTGCTCGAGAAATCTGTC 657
Db 651 GTTATCAGCTAGCAGTGTGCTGCTGAGAGCGCTTATGGAATGG----- 695
Qy 658 AAATTCAGAGACCTATCGCGACAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACC 717
Db 696 -GAAGCAGGGATCTGCTCTTCTTCAAGTTCAGTTCAGTAAATGCAAAATCCAGCTCAACCAA 754
Qy 718 TTCTGTAAAGCAATCCAGGAACCTACAAATATAACACAAGCAACACAGACACACTAACAG 777
Db 755 GATTGTGAGAGAACCAACAATCT-----GACTGTGCTACTCCAACCTTA 794
Qy 778 TCAAGAAAGTCAAAAGTAAATGACATGATAGTCTCTTGAACAGCTCTACAAACTGCCCTTT 837
Db 795 TCATCAAAATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTT 854
Qy 838 GAGTCAACGACATGTAGAAATCTGATGGCTTTGTCTTTGATCCAGACAAAATCACAAGTCG 897
Db 855 ATCAGAACCGCATGTGGAAATCTGATGGCTTTATTTTCGACCCAGCGCAAAATCACAAGTCG 914
Qy 898 AACAGCTAGAGTGTTCAGTGTCCACACGAGATCATTAACACTTCATCCCTTACTCTCA 957
Db 915 AACGCCAGAGGTGTAGTGTCTCCCTCATGTGTAAACCATTAACCACTTTATCCCTTATGAACA 974
Qy 958 AATGCTGAATTTGAGAAAGCAAGTCTGCTGATTTATTCCTTCTGTTATCGTTCAAACCA 1017
Db 975 AATGCTGAATTTGAGAAAGCAAGTCTGCTGATTTATTCCTTCTGTTATCGTTCAAACCA 1034
Qy 1018 TTGGGTACAGATTCAAGGCCAGAAACAACCAAGTCCCAACCGACTCCGGAAACCTAGTCC 1077
Db 1035 TTGGGTACAGATTCAAGGCCAGAAAGCAAGTCCCAACCGACTCCAGAAACCTAGTCC 1094
Qy 1078 AGGCCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTCTTGTGTTAGTCAGCT 1137
Db 1095 AAGTCCGCAACCCAGCTCCAAAGCAATCCAAATTTGATGGGA-----ATTGGTCAAAGAAC 1148
Qy 1138 GGTACGAAAGTTGGGAGGATATGTTTCCAAAGAAAAGGCAATCTCTGCTGTTATGCTTT 1197
Db 1149 TGTTCGAAAGTGGGATGGTATGTTTCTTTCAGGAGATGAGTTTCTCGTTATATCCC 1208
Qy 1198 TCGAAAGATTTTACCATCTGAAACTGTTTAAATCTTTGAAAGCAAGTATCAAAACAAGA 1257
Db 1209 AGCCAGGATCTTTCAGCAGAAACACAGCAGCAGCATTTGATAGCAAACTGCCCAAGCAGA 1268
Qy 1258 GAGTGTTCACACATTTTAACTGCTTAAAGAAAATGTTGCTCTCGTGCACCAAGAAAT 1317
Db 1269 AAGTGTATCTCTAAGCTAGGAACCTAAGAAAACCTGACCTCCCATCTAGTATCGAGAT 1328
Qy 1318 TTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAAATAGGG 1377

Db	1329	TTACANTAGGCTTATGACTTACTACGAGAATTCCACAAGATTACTTGTATAAAGG	1388
QY	1378	TGCTAATCTGATTTCCAAAGCCTTACAGCAAAATTATTAGAACGCTTGAATGATGAATCGAC	1437
Db	1389	TCGACAAGTTGATTTTGAGGCTTTTGATATAAAGCTGTTGGAACGACTCAAGAGTGTCTCAAG	1448
QY	1438	TAAATAAGAAAATTTGGTTAGATGATTTATGCGCAATCTAGCACCAATTACCATCCAGA	1497
Db	1449	TGATAAAGTCAAGTTAGTGGAGATATTCTTGCTTCTTAGCTCCGATTTGTCATCCAGA	1508
QY	1498	GGGACTTGGCAAAACCAAAATCTCAAAATGAGTATACCTAGAGACCAAGTTTGGTATGCTCA	1557
Db	1509	AGTTTAGGAACCAATGGCAATTTACCTACACTGATGATGAGATTCAAGTAGCCAA	1568
QY	1558	ATTAGCTGATTAAGTATACAAGCTCAGATGGTTACATTTTGTGAACATGATATAATACG	1617
Db	1569	GTTCGACGCAAGTACACAGCAGAAGACGGTTATATCTTTGATCCTCGTATATACCAG	1628
QY	1618	TGATCAAGGAGATGCATATGTAAAGCCTCATATGGGCCATAGTCACCTGGATTGGAAAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTGGATTAAAAAGA	1688
QY	1678	TAGCCTTTCTCATAGGAAAAAGTTGCAGCTCAGCCCTATACCTAAAGAAAAGGTATCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGCGGAGCCCGAGGCTTATGCTTAAAGAGAAAGGTTTGAC	1748
QY	1738	ACCTCCATCTCCAGCCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACGACTAT	1797
Db	1749	CCCTCCTTCGACACACCATCAGGATTCAGGAATCTGAGGCAAAAGGACGAGACTAT	1808
QY	1798	TTCAATCTGTGAAAGGGGAAAAAGCAATCCACTCGTTTCGACTTCCATATATGTTGCA	1857
Db	1809	CTACAAACGCGTGAAGACAGCTAAGAGGTGCCACTGTATCGTATGCTTACAACTTCA	1868
QY	1858	GCATACAGTTGAGGTTAAAAACGTAATTTGATTATTCCTCATAGGATCATTTACCAAT	1917
Db	1869	ATATACTGTAGAAGTCAAAACCGGTAGTTTAATCATACCTCATTTAGCCATTACCAT	1928
QY	1918	TATTAAATTTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGA	1977
Db	1929	CATCAAAATTTGAGTGGTTTTCGCAAGGCCTTATGAGGCACCTAAGGGGTATACTCTGA	1988
QY	1978	AGATTTGTTTGGCAAGATTAAGTACTAGTGAACAACCTGTCGCAAGCTCCCATCTTAA	2037
Db	1989	GGATCTTTTGGCACTGTCAAGTACTATGTGCAACATCCAAACGAAAGCTCCGCAATC	2048
QY	2038	TGATGATGGGCAATGGCCAGTCAGCATGTGTTAGCCAGAAGACCCAGTGAAGATCC	2097
Db	2049	TANTGGTTTGGTTAAAGCTTAGCAGACCATGTTCAAGAAAACAAAATGGTCAAGCTGAT	2108
QY	2098	AAATAAGAACTTCAAGCGGATGA	2121
Db	2109	CAATCAAAACGAAAAACCAAGGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAC	2168
QY	2122	-----AGAGCCAGT	2130
Db	2169	CCCTCGAGAAGAAACCGCAAGCGAGAAAAACAGAGTCTCCAAAACCAACAGAGAAC	2228
QY	2131	AGAGAAAAACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAGTAGAAGCCCA	2190
Db	2229	AGAAGATCACAGAGGAATCAGAAGACCTCAGGTCGAGACTGAAAAGGTTGAAGAAA	2288
QY	2191	ACTCAAGAGCAGAAGTTTCTTCCGAAAGTAACGGATTCAGTCTGAAAGCCAAATGC	2250
Db	2289	ACTGAGAGAGGCTGAAGATTTACTTTGGAAAAATCCAGGATCCAATATCAAGTCAATGC	2348
QY	2251	ACAGAAAACTTACGTGTTTACGAAATAATTTGACTCTTCAAAATTATGGATAACAATAG	2310
Db	2349	CAAGAGACTCTCACAGGATTTAAAAATAAATTTACTATTTCGCCCCAGGACACAATAC	2408
QY	2311	TATCATGGCAGAGCAGAAAAATTTACTTCGTTGTTTAAAGAGATTA	2358

Db 2409 TTATTATGGCAGAAAGCTGAAAAAATACTATTGGCTTTTATTAAAGAGAGTAA 2456

RESULT 11
US-10-387-783-11
; Sequence 11, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-11

Query Match 41.5%; Score 991; DB 16; Length 2531;
Best Local Similarity 65.7%; Pred. No. 8.5e-230;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAAACGGTTAAGGAAA---TAATCGTGTTC	57
Db	60	TTCTATAGAGTTGGAGCTTACCAAGCTGGTCAGATTAAGAAAGATCTAATCGAGTTGC	119
Qy	58	CTATATAGATGGAAAAACAAGCGACGCAAAAAACGAGAAATTGACTCTCTGATGAGTTAG	117
Db	120	TTATATAGATGGTGTATCAGGCTGGTCAAAAGGCGAAAAACTTGACACCAAGATGAAGTCAG	179
Qy	118	CAAGCGTGAAGGATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGT	177
Db	180	TAAAGAGGAGGGATCAACGCCAACAAATGGTTATCAAGATTAACGATCAAGGTTATGT	239
Qy	178	CACCTTCATCGGCGACCCTATCATTTATTAACAATGGTAAGGTTCTTTATGACGCTATCAT	237
Db	240	GACCTCTCATGAGACCAATTATCATTACTATATATGCAAGGTTCTTTATGATGCCATCAT	299
Qy	238	CAGTGAAGAATTACTCATGAAGATCAAACTATAAGCTAAAAAGATGAGGATATTGTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAGTTCAAGGATTCACACATTGTCAA	359
Qy	298	TGAGGTCGAAGGTGGATATGTTATCAAGTAGATGAAAATACTATGTTTACCTTAAGGA	357
Db	360	TGAATCAAGGTGGTTATGTCAATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA	419
Qy	358	TGCTGCCCGCGGATAACGTCGTTCAAAAGAGGAAATCAATCGACAAAAACAAGCA	417
Db	420	TGCRGCTCATCGGATAATATTCGGCAAAAGNAGAGATTAACGTCAGNAGCAGGAACG	479
Qy	418	TAGTCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGTTGCCCTTGGCACGTTTC	477
Db	480	CAGTCAATAATCAT-----AACTCAAGAGCAGATAATGCTGTGTGCGACCGAGGC	530
Qy	478	GCAAGGACGCTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGA	537
Db	531	CCAAGACGTTATACACGGATGATGGGTATATCTTCAATGATCTGTATCATTTGAGGA	590
Qy	538	TACTGGTGATGCTTATATCGTTCCCTCATGGAGATCAATACCAATTACATTCCTTAAGNATCA	597
Db	591	CACGGGTGATGCTTATATCGTTCTCTCAACGGCGACCATTAACCAATTAATTCCTTAAGATGA	650
Qy	598	GTTATACGTACGAGTTGGTGTGCGAAGACCTTCCTATCTGTGCGAGAAATCTGTTC	657

```

1689 TAGTTTGTCTGAAGCTGAGAGCGCGAGCCAGCCAGGCTTATGCTRAAGAGAAAGTTTGAC 1748
1738 ACCTCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGAGATGAGTAGTGACAGCTAT 1797
1749 CCCTCTTCCAGACAGCATCAGGATTCAGGAAATCTAGGCAAAAGAGGAGCAGAGCTAT 1808
1798 TTAACAATCTGTGTAAGAGGGAAGAAACGAATTCACCTCGTTCCGACTTCCATATATGGTTGA 1857
1809 CTACAAACCGGTGTAAGAGCGCTAAGAGGTCACCTTGAATGCTATGCTTCAATCTTCA 1868
1858 GCATACAGTTTGTAGTTAAAGCGTAAATTTGATTTATTCCTCATAAAGGATCATACCAATAA 1917
1869 ATATACCTGTGAAGAGTCAAAACCGTAGTTTAAATCAATACCTCATTTATGACCATTAACCAATA 1928
1918 TATTAAATTTGCTTTGGTTTGTATGATCAACATCAAAAGCTCCAAATGGCTATATACCTTTGA 1977
1929 CATCAATTTTGTGTTTTCAGCAAGGCTTTTATAGGCGACCTAAGGGGTATATCTTTGA 1988
1978 AGATTTGTTTGGACGATTAAGTACTAGTACGAGACCTGACGAGCTGACACATCTCTAA 2037
1989 GGATCTTTTGGCGACTGTCAGTACTATGTCGAACATCCAAACGAGCTCGGATTCAGA 2048
2038 TGATGATGGGGCAATGCCAGTGAAGCTGTGTTAGGCAAGAAAGACCAACAGTGAAGATCC 2097
2049 TAATGGTTTGTGTAACGCTAGGACCATGTTCAAAAGAAACAAATAATGCTCAAGCTGATAC 2108
2098 AAATAAGAACTTCAAGCGGATGA----- 2121
2109 CAATCAAGCGAAACCAAGCGAGGAGAAACCTCAGACAGAAAAAOCCTGAGGAGAAAC 2168
2122 -----AGAGCCAGT 2130
2169 CCCTCGAAGAGAGAAACCGCAAGCGAGAACACAGAGTCTCCAAACCAACAGAGAAC 2228
2131 AGAGAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCA 2190
2229 AGAAGATCACAGAGGATCAGAGAACCTCAGTGCAGACTGAAAAAGTTGAAGAAA 2288
2191 ACTCAAGAAAGCAGAGTTTGTCTGCGAAAGTAACGATTCCTAGTCTGAAAGCCAATGC 2250
2289 ACTGAGAGAGCTGAAGATTTACTTTGGAATAATCCAGATCCATATCAAGTCCATGC 2348
2251 AACGAAACTCTAGCTGCTTACGAAATAATTTGACTCTTCAATATATGGATAACAATAG 2310
2349 CAAGAGAGCTCTCAGAGATTAATAAATAATTTACTATTTGSCACCCAGGACACAATAC 2408
2311 TATCATGGCAGAGCAGAGAAATTTACTTTGCGTTGTTAAAGGAAGTAA 2358
2409 TATTATGGCAGAGCTGAAATAATTTGCTTTTAAAGGAGAGTAA 2456

RESULT 12
US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA

```

; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

Query Match	41.4%	Score 990;	DB 13;	Length 2531;
Best Local Similarity	68.1%	Pred. No. 1.5e-229;		
Matches 1478; Conservative	0;	Mismatches 636;	Indels 57	

QY	1	TTCTTACGAGTTGGACGTGTATCAAGCTAGAACCGTTAAGGAAA---TAATCGTGTTTC	57
Db	60	TTCCATATGAACTTGGTCGTACCAACAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC	119
QY	58	CTATATAGATTGGAATAACAAGCGACGCAAAACCGAGNAATTTGACTCTCTGATGAGGTTAG	117
Db	120	TTATATAGATTGGTCATCAGGCTGTCAAAGCGCAAAACCTTGACACACAGATGAAGTCAG	179
QY	118	CAAGCGTTGAAGGAATCAATCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGT	177
Db	180	TAAGAGGAGGAGGATCAACGCGCAACAAATCGTCAATCAAGATTACGGATCAAGGTTATGT	239
QY	178	CACCTTCATGCGGACCACTATCATTTATACAATGGTAAAGTTCCTTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGACCATTTATCATTTACTATTAATGGCAAGGTCCTTTATGATGCCATCAT	299
QY	238	CAGTGAAGAATTACTCATGAAGAATCCAAACTATAAGCTAAAGATCAGGATATTGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTATCAGTTGAAGGATTCAGACATTTGTCAA	359
QY	298	TGAGGTTCAAGGTTGGATATGTTTATCAAGCTTAGATGGAAATATCTATGTTTACCTTAAGGA	357
Db	360	TGAAATCAAGGTTGGTTATGTTTATCAAGCTTAGATGGAAATATCTATGTTTACCTTAAGGA	419
QY	358	TGCTGCCACGCGGATAACGTCGCTCAAAAGAGGAAATCAATCGACAAACAAAGACGA	417
Db	420	TGCAGCTCATGCGGATAATATTCGGACAAAGAGAGATTTAAAGTTCAGAAAGCAAGCA	479
QY	418	TAGTCAACATCTGTAAGGTTGGAACTCAAGAAAGATGGTGTGTGTGCTTGGCATTGAGG	477
Db	480	CAGTCAATATCAAGGGGTGGTTCT-----AAGCATCAAGCAGTAGTTGCAGCCAGAGC	533
QY	478	GCAAGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGA	537
Db	534	CCAAGCAGCTATACAAACGATGATGGTTATATCTTCAATGATCTGATATCATTTGAGGA	593
QY	538	TACTGTTGATGCTTATATCGTTTCTCATGGAGATCAATTACCATTCATTCCTTAAGAATGA	597
Db	594	CACGGTGTATGCTTATATCGTTCTCTCAGCGGACCAATTACCATTCATTCCTTAAGAATGA	653
QY	598	GTTATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCTCTATCTGTGTCGAGGAATCTGTC	657
Db	654	GTTATCAGCTAGCGAGTTAGCTGCTGCAGAGCCT-----	688
QY	658	AAATTCAGAACCTATCGCCGCAAAATAGCGATAACATTCAGAGCAACAACTGGGTACC	717
Db	689	-----ATTGGAATGGGAAGCAGGAGTCTGTCTCTTCTTCAAGTTCTAGTTTATAATGC	740
QY	718	TTCTGTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAG	777
Db	741	AAATCCAGCTCAACCAAGATTGTCAGAGAACCAAACTGACTGTCACTTCCAACTTATCA	800
QY	778	TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTTGAAACAGCTCTACAAACCTGCTTT	837
Db	801	TCA---AAATCAAGGGGAAAAATTTCAAGCCTTTTACGTGAATGTTATGCTTAACCCCTT	857
QY	838	GAGTCAACGACATGTAGAAATCTGATGGCCCTGTCTTTGATCAGACCAAAATCAAGTCG	897
Db	858	ATCAGACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGGCGCAATCACAGTCG	917
QY	898	AACAGCTAGAGGTGTGCAAGTCCCAACAGGAGATCAATTACCACTTTCATCCCTTACTCTCA	957
Db	918	AACGCCAGAGGTGTAGCTGTCCCTCATGGTAAACCATTACCACTTTATCCCTTATGAACA	977
QY	958	AATGTCGAATTGGAAAGCAAGATCGTCGTATTAATTCCTCTCGTTATCTGTTTCAAAACA	1017

Db	978	AAATGCTGAAATTGGAAAAACGAATTTGCTGTAATTTATTTCCCTTGCTTATCGTTTCAAAACCA	1037
Qy	1018	TTGGGTACACAGATTCAGAGCCAGAACCAAGTCCACACCGGACTCCGNAACCTAGTCC	1077
Db	1038	TTGGGTACAGATTCAGACAGAACCAAGTCCCAATTCGACTCCGGAACCTAGTCC	1097
Qy	1078	AGGCCCGAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTTC-----TTT	1125
Db	1098	AAAGTCGCAACCTGCACCAAAATCTCAACAGCTCCAAGCAATCCAATTCATGAGAAAT	1157
Qy	1126	GGTTAGTCAGCTGGTACGAAAGTTGGGGAAGGATATGTATTCCAGAAAGAGGCATCTC	1185
Db	1158	GGTCAAGAAGCTGTTCGAAAGTAGCGATGGTATGTCTTTGAGGAGAATGGAGTTTC	1217
Qy	1186	TCGTTATGCTTTGCGAAGATTTACCATCTGAAACGTTTAAAACTTTGAAAGCAAGTT	1245
Db	1218	TCGTTATCCAGCCAGGATCTTTCAGCAGAAACAGCAGGAGGATTTGATAGCAAACT	1277
Qy	1246	ATCAAAACAGAGAGTGTTCACACACTTTAACTGTCTAAAAAGAAAAATGTTGCTCCTCG	1305
Db	1278	GGCCAAAGCAGGAAAGTTATTTCTCATAGCTAGGAGCTAGAAAACTGACCTCCCATCTAG	1337
Qy	1306	TGACCAAGAATTTTATGATAAGCATATATCTGTGTTAACTGAGGCTCATAAAGCCTTGT	1365
Db	1338	TGATCGAGAAATTTTACAATAAGGCTATGACTTTACTAGCAAGAATTCACCAAGATTTACT	1397
Qy	1366	TGNAATAAGGGTCGTAAATCTGAATTTCCAAGCCTTTAGCAAAATTTATAGAAGCTTGAA	1425
Db	1398	TGATAATAAGGTCGACAGATTTGATTTGAGGCTTTGGATAAGCCTGTTGGAACGACTCAA	1457
Qy	1426	TGATGAATTCGACTAATAAAGAAAAATTTGTTAGATGATTTATTTGGCAATCTCTAGCACC	1485
Db	1458	GGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATCTTGCCCTCTTAGCTCCGAT	1517
Qy	1486	TACCATTCAGAGGACTTCGGAAAAACAAATTTCTCAAAATGAGTATPACTGAAGACGAAGT	1545
Db	1518	TCGTCATCCAGAACGTTTGTAGAAAAACCAATGCGCAAAATTAAGCTACACTGATGATGAT	1577
Qy	1546	TCGTATTGCTCAATTAGCTGATAAGTATCAAGTCCAGATGGTTACATTTTGTATGAACA	1605
Db	1578	TCAGTAGCCAGTTTGGCAGGGAAGTACAAACAGAAAGACGGTTATATCTTTGATCCTCG	1637
Qy	1606	TGATPATAATCAGTGATGAAGGAGATGCATGTATGACGCTCATATGGGCCATAGTCACTG	1665
Db	1638	TGATATAACCAAGTGATGAGGGGATCCCTATGTAATCCACATATGACCCATAGCCACTG	1697
Qy	1666	GATTTGGAAGATAGGCTTTCTGATAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA	1725
Db	1698	GATTTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCTAAAGA	1757
Qy	1726	AAAAGGTATCCTACCTCCATCTCCAGAGCAGATGTTAAAGCAATCCAACTGGAGATAG	1785
Db	1758	GAAAGGTTTGACCCCTTCCTTCACAGACCATCAGGATTCAGAAATATCTGAGGCCAAAGG	1817
Qy	1786	TGCAGCAGCTATTTACAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCC	1845
Db	1818	AGCAGAAGCTATCTACAACCGGTGAAGAGCAGCTAAGAAGGTGCCACTTGATCTGATGCC	1877
Qy	1846	ATATATGTTGAGCATACAGTTGAGGTTTAAAAACGGTAAATTTGATTTCTCTCAAGGA	1905
Db	1878	TTACAATCTTCAATACTGTAGAAAGTCAAAAACGGTAGTTTAATCATACCTCATTTATGA	1937
Qy	1906	TCATTACCATAATTTAAATTTGCTTTGATGATCACACATACAAGCTCCAAATGG	1965
Db	1938	CCATTACCATAACATAAATTTGAGTTGTTGACGAAGGCCCTTTATGAGGCACCTCAAGGG	1997
Qy	1966	CTATACCTTTGGAAGATTTGTTTGGCACGATTTAAGTACTACGTAGAACACCTCGAGAACG	2025
Db	1998	GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAACG	2057
Qy	2026	TCCACATCTTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAGACCA	2085
Db	2058	TCGCATTCAGATAATGTTTGTGTAACGCTTAGCGACCATGTCGTATTAATATGATGATG	2117

QY 1606 TGATATATCAAGTATGAAGAGATGATATATGAAGCGCTCATATGGCCATAGTCACTG 1665
 Db |||||
 QY 1638 TGATATCAAGTATGAAGAGATGATATATGAAGCGCTCATATGGCCATAGTCACTG 1697
 Db |||||
 QY 1666 GATTGGAAGATAGCTTCTGATGAAGAAAGTTGAGCTCAAGCTTAAAGA 1725
 Db |||||
 QY 1698 GATTGGAAGATAGCTTCTGATGAAGAAAGTTGAGCTCAAGCTTAAAGA 1757
 Db |||||
 QY 1726 AAAAGTATCCCTACCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTGGAGATG 1785
 Db |||||
 QY 1758 GAAAGTATCCCTACCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTGGAGATG 1817
 Db |||||
 QY 1786 TGCAGCAGCTATTAACAATCGTGTGAAAGGGGAAAGAAAGCAATCCAACTGGAGATG 1845
 Db |||||
 QY 1818 AGCAGAAGCTATCAACAACCGGTGAAAGCAGCTAAAGAGGTGCCACTTGATGCC 1877
 Db |||||
 QY 1846 ATATATGTTGAGGATACAGTTGAGTTAAAGCGGTAAATTTGATTTCCCTCAATAGGA 1905
 Db |||||
 QY 1878 TTCAATCTTCAATATATCTAGAGAGTCAAAAGCGGTAAATTTGATTTCCCTCAATAGGA 1937
 Db |||||
 QY 1906 TCATTACCATTAATTAATTTGTTGCGAGCAATTAAGTACTACGTAGAACACCTTGACGAGC 2025
 Db |||||
 QY 1938 CCATTACCATTAATTAATTTGTTGCGAGCAATTAAGTACTACGTAGAACACCTTGACGAGC 2057
 Db |||||
 QY 1966 CTATACCTTGAAGATTTGTTGCGAGCAATTAAGTACTACGTAGAACACCTTGACGAGC 2085
 Db |||||
 QY 2026 TCCACATTTCAATGATGAGTGGGCAATGCCAGTACGATGTTAGGCAAGAGACCA 2117
 Db |||||
 QY 2058 TCCGATTCAGATTAATGTTTGTGTAAGCTAGGACCATTTCCGTAAGAGTAAAGTAA 2145
 Db |||||
 QY 2086 CAGTGAAGATCCAAATGAAGACTTCAAGCGGATGAAGAGCCAGTGAAGTAAAGTAAAGTAA 2177
 Db |||||
 QY 2118 CCAAGACAGTAAACCTGATGAAGATGAAGACATGATGAAGTAAAGTAAAGTAAAGTAA 2177
 Db |||||
 QY 2146 TGAGCCAGAAG 2156
 Db |||||
 QY 2178 TGAATCTGATG 2188
 Db |||||

RESULT 14
 US-10-387-783-5
 ; Sequence 5, Application US/10387783
 ; Publication No. US20040005331A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; FILE REFERENCE: 469201-683
 ; CURRENT FILING DATE: 2003-03-13
 ; PRIOR FILING DATE: 1998-12-21
 ; PRIOR FILING DATE: 1998-12-21
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patent In Ver. 3.0
 ; SEQ ID NO 5
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; US-10-387-783-5

Query Match 41.4%; Score 990; DB 16; Length 2531;
 Best Local Similarity 68.1%; Pred. No. 1.5e-229;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;
 QY 1 TTCTTACAGTGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC 57
 |||||

Db 60 TTCCTATGAACCTTGGTCTGTCACCAAGCTGGTAAAGAAAGAGCTTAATCGAGTTTC 119
 QY 58 CTATATAGATGAAACAAAGCGAGCGCAAAACCGAGAAATTTGACTCTCTGATGAGGTAG 117
 Db |||||
 QY 120 TTATATAGATGATGATGAGCGTGGTCAAAAGCGAGAAACCTTGACACCGAGATGAAGTCAG 179
 Db |||||
 QY 118 CAAAGCTGAGGAATCAATTCGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATCTG 177
 Db |||||
 QY 180 TAAAGGAGGAGGAGTCAACCGCGAAACAAATCGTCAATCAAGATTCGGATCAAGGTATATG 239
 QY 178 CACTTCACATGCGACCACTATCATTTATCAATGGTAAAGTTCCTTATGACGCTATCAT 237
 Db |||||
 QY 240 GACCTCTCATGAGACCACTATCATTTACTATAATGCGCAAGGTCCTTATGATGCCATCAT 299
 QY 238 CAGTGAAGAAATTTACTTCATGAAAGATCCAACTATTAAGCTTAAAGATGAGGATTTGTTAA 297
 Db |||||
 QY 300 CAGTGAAGAGCTCTCTCATGAAGATCCGAAATTCAGTTGAAGGATTCAGACATTTGTCAA 359
 QY 298 TGAGTCAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGCA 357
 Db |||||
 QY 360 TGAATCAAGGCTGATTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGCA 419
 QY 358 TGCTGCCACCGCATACGTCCTGTAACAAAGAGAAATCAATCGACAAAACAAAGAGCA 417
 Db |||||
 QY 420 TGCAGCTCATGCGGATTAATTCGACCAAAAGAGAGATTAAGGCTCAGAAAGCAGGAACA 479
 QY 418 TAGTCAACATCGTGAAGGTGAACTCCAAGAAACGATGGTCTGTCCTGTCGACGCTTC 477
 Db |||||
 QY 480 CAGTCAATATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 533
 QY 478 GCAAGGAGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGCA 537
 Db |||||
 QY 534 CCAAGGAGCTTATACAAAGGATGATGTTATATCTTCAATGCACTCATATCATAGAGCA 593
 QY 538 TACTGGTATGCTTATATCGTTCCTCATGAGATCATTTACATTCATTCCTTAAGAAATGA 597
 Db |||||
 QY 594 CACGGTATGCTTATATCGTTCCTCAGGCGACCATTTACATTTACATTCCTTAAGAAATGA 653
 QY 598 GTTATCAGCTAGCGAGTTGGTCTGCTGAGAAAGCTTCTTATCTGGTTCGAGAAATCTGTC 657
 Db |||||
 QY 654 GTTATCAGCTAGCGAGTTAGTCTGCTGAGAGCT----- 688
 QY 658 AATTCAGAACTTATGCGCGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACC 717
 Db |||||
 QY 689 -----ATTGGAATGGGAAGCAGGATCTCGTCTCTTCTCAAGTTCTAGTTATATGTC 740
 QY 718 TTCTGTAAGCAATCCAGGAACTACAAATACATAACAGCAACCACTGACTGTCTCAACTTATCA 800
 Db |||||
 QY 741 AATTCAGCTCAACCAAGATTTGTCAGAGAACCACTGACTGTCTCAACTTATCA 800
 QY 778 TCAAGCAAGTCAAGATTAATGACATTTGATGATGCTCTTTGAAACAGCTCTCAAACTGCTTT 837
 Db |||||
 QY 801 TCA---AATCAAGGGAACAACTTTCAAGCTTTTACGTGAATTTGATGCTTAAACCTTT 857
 QY 838 GAGTCAACGACATGATAGATTCGATGAGCTTCTTTGATCCAGCACAAATCACAGTCG 897
 Db |||||
 QY 859 ATCAGAACCCCATGTGGAATCTGATGGCTTATTTTGAACCCAGCGCAATCACAGTCG 917
 QY 898 AACAGCTAGAGTGTGAGTGCACACGAGATCATTTACCACTTCCTTCTCTCTCA 957
 Db |||||
 QY 918 AACCCGAGAGTGTAGCTGCTCCCTCATGGTAAACCATTTACCACTTTATCCCTTATGAACA 977
 QY 958 AATGCTGAATTTGAAGAACGAATCGCTCGTATTTATCCCTTCTGTTATCGTTCGTTCAACA 1017
 Db |||||
 QY 978 AATGCTGAATTTGAAGAACGAATTTGCTGCTATTTATCCCTTCTGTTATCGTTCGTTCAACA 1037
 QY 1018 TTGGTACAGATTTCAAGGCGAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCC 1077
 Db |||||
 QY 1038 TTGGTACAGATTTCAAGGCGAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCC 1097
 QY 1078 AGGCCGCAACCTCGACCAATCTTAAATGAGCTCAAAATCTTTC-----TTT 1125
 Db |||||
 QY 1098 AAGTCCGCAACCTCGACCAATCTTCAACCGACTCCAGCAATCAATTTGATGAGAAAT 1157
 Db |||||

1126 GGTAGTACGCTGGTAAAGCTGGGGAAGGATATGTATCGAAGAAAGGCACTTC 1185
 1158 GGTCAAGAAGCTGTTGCAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTC 1217
 1186 TCGTTATGCTTTGCGGAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTT 1245
 1218 TCGTTATATCCAGCCAGGATCTTTTCAGCAAGAACAGCAGCAGGCAATGATGCAAACT 1277
 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCG 1305
 1278 GGCAAGCAGGAAAGTTTATCTATCACTAGTAGAGCTTAAGAAAGTACCTCCCATCTAG 1337
 1306 TGACCAAGATTTTATGATAAGCATATAATCTGTTAACTGAGGCTGTAAGAGCTTTGTT 1365
 1338 TGATCGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
 1366 TGNAAATGAAGGCTGTTATCTGATTTCCAGAGCTTACAGCAAAATTTATGAAAGCTTGAA 1425
 1398 TGATAATAAGAGTGCAGCAAGTTGATTTTGAGGCTTTGGATTAACCTGTTGGACGACTCAA 1457
 1426 TGATCAATCGACTTAATAAGAAATTTGCTAGATGATTTATTTGGCAATTCCTAGCACCAAT 1485
 1458 GGATGTCCTCAAGTGAATAAGTCAAGTTAGTGAATGATATCTTGCTTCTTAGCTCCGAT 1517
 1486 TACCATCCAGAGGCTTGGCAACCAAAATTTCTCAAAATGAGTATCTAGAGCAAGT 1545
 1518 TCGTCATCCAGAACGTTTAGGAAACCAAAATTCGCAAAATTCCTACACTGATGATGAT 1577
 1546 TCGTATGCTCAATAGCTGATAGTATACACGTCAGATGCTTACATTTTGTGTAACA 1605
 1578 TCAAGTAGCAAGTTGGCAGGCAAGTACACAGAGAGGTTATCTTTGATCTCTCG 1637
 1606 TGATATAATCACTGATGAAGGAGATGCATATGATTAAGCGCTCATATGGCCATATGACTG 1665
 1638 TGATATAACCACTGATGAGGGGATGCTTATGTAATCCACATATGACCCATAGCCACTG 1697
 1666 GATTGGAAGATGAGCTTTCTGATGAAGAAAGTTGAGCTCAAGCTTATCTAATAACA 1725
 1698 GATTAAAGAAAGTAGTTTCTGTAAGCTGAGAGCGGCGAGCCAGGCTTATGCTAAGA 1757
 1726 AAAAGTATCCTACTCCATCTCCAGCAGCAGATGTTAAAGCAAAATCCAACTGGAGATAG 1785
 1758 GAAAGCTTTGACCCCTCTTCCAGACACCATCAGGATTCAGGAAATCTGAGCAAAAG 1817
 1786 TGACGAGCTATTTCAATCGTGTGAAGGGGAAAGCAATTCCTACTGCTTCACTTCC 1845
 1818 AGCAGAGCTATCTCAACCGCTGAAAGCAGCTAAGAGGTCGCCACTTGATCGTATGCC 1877
 1846 ATATATGCTTGAACATACAGTTGAGGTTTAAAGCGTTAATTTGATTTATCTCTCATAGA 1905
 1878 TTCAATCTTCAATATCTAGAGGTCAAAGACGGTAGTTTAACTCATATCTCATATGA 1937
 1906 TCATTACCAATATTAATTTGCTGTTGATGATCACATACAAAGCTCCAAATGG 1965
 1938 CCATTACCAATCAATCAATTTGAGTGGTTTGACGAGGCTTTATGAGGCACTTAAGG 1997
 1966 CTATACCTTGAAGATTTGTTTGGCAGCTTAAGTACTTACGTGAACACCTTGAAGAG 2025
 1998 GTATCTCTTGAAGATCTTTTGGCGCTGTCAGTACTATGTCGAACATCCAAACGAA 2057
 2026 TCCATCTTGAAGTGGGCAATGCCAGTGAGCATGTTTGGTAAAGTGGTGGTAAAGAGCA 2085
 2058 TCCGATTCAGATAATGTTTGGTAAAGTGGTGGTAAAGTGGTAAAGTGGTAAAGTGGT 2117
 2086 CAGTGAAGATCCAAATGAAGACTTCAAGCGGATGAAGAGCCAGTGAAGAGCAACCTGC 2145
 2118 CCAAGACAGTAACTGATGAAGATAGGAATGATGAAGTGAAGTGAAGTGAAGTGAAG 2177
 2146 TGAGCCAGAG 2156
 2178 TGAATCTGATG 2188

RESULT 15
 US-09-765-272-65
 ; Sequence 65, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 65:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2290 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
 ; US-09-765-272-65
 Query Match 41.3%; Score 987.6; DB 9; Length 2290;
 Best Local Similarity 67.7%; Pred. No. 5.3e-229;
 Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
 QY 1 TTCTTACGAGTTGGGCTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC 57
 DB 4 TTCTTACGAGTTGGGCTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC 63
 QY 58 CTATATAGATGGAAGAAACAAAGCGACGCAAAAGAAATTTGACTCTCTGATGAGTTAG 117
 DB 64 TTATATAGATGAGTATGATCGCTGCTCAAGAGGAGGAAATCTTGACACGATGAAGTCAG 123
 QY 118 CAAGCGCTGAGGAAATCAATGCTGAGCAAAATCGTCTCATCAAGATTAACAGCAAGCTATGT 177
 DB 124 TAAGAGGAGGAGGATCAACCGCAAAATNGTATCAAGATTAACGATCAAGGTTAATGT 183
 QY 178 CACTTCACTGAGGAGGATCAATGCTGAGCAAAATCGTCTCATCAAGATTAACAGCAAGCTATGT 237
 DB 184 GACCTCTCATGAGGAGGATCAATGCTGAGCAAAATCGTCTCATCAAGATTAACAGCAAGCTATGT 243
 QY 238 CAGTGAAGAAATTAATCTCATGAAAGATCAAACTATAAGCTAAAGATGAGGATATGTATA 297
 DB 244 CAGTGAAGAGCTCTCTCATGAAAGATCGAATTAATCAAGGATTAAGGATTAAGGATTAAG 303
 QY 298 TGAGGTCAAGGCTGATGATTTATCAAGGTAGATGGAAGAAATCTATGTTACCTTAAGGA 357
 DB 304 TGAATCAAGGCTGATGATTTATCAAGGTAGATGGAAGAAATCTATGTTACCTTAAGGA 363
 QY 358 TGCTGCCCGCGGATTAAGCTCGCTGACAAAGAGGAAATCAATCGACAAAGAGCA 417

Db 364 TGCAGCTCATGGGATAATATTCGGACAAAGAGATTAACGCTCAGAGCAGGACG 423
QY 418 TAGTCAACATCGTGAAGCTGGAACCTCCAGAAACGATGGTGTCTGTTCGCTTGGACGATTC 477
Db 424 CAGTCATAATCAT-----AATCAAGACGACGATTAATGCTGTGTGCGCAGCAGACG 474
QY 478 GCAAGGACGCTATACACAGATGATGGTGTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db 475 CCAAGGACGTTATACACGATGATGGGTATATCTTCAATGCACTGATATCATTTGAGGA 534
QY 538 TACTGGTATGCTTATATCGTTCCTCATGAGATCAATACCATACATTCCTTAAGATGA 597
Db 535 CACGGGTGATGCTTATATCTGTTCTCTCAGCGGACCATTACCATTACATTCCTTAAGATGA 594
QY 598 GTTATACAGTACGAGTGGCTGTGTCAGAGACCTTCTATCTGGTCGAGGAATCTGTC 657
Db 595 GTTATACAGTACGAGTGGCTGTGTCAGAGACCT----- 629
QY 658 AATTCAGAACCTATCGCCGACAAATAGCGATAACACATTCAGAAACAACTGGGTACC 717
Db 630 -----ATTGGAATGGAGCAGGATCTCGTCTCTTCTCAAGTTCCTAGTTAATATGC 681
QY 718 TTCTGTAAGCAATCCAGGACCTACAATATTAACAAGCAACCAACGACCACTAACAG 777
Db 682 AATTCAGCTCAACCAAGATTTGTGAGAGAACCAATCTGACTCTCACTCCAACTTATCA 741
QY 778 TCAAGCAAGTCAAAAGTAAATGACATTTGATGCTCTCTTGAACAGCTCTACAAACTGCCTTT 837
Db 742 TCA--AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCTT 798
QY 838 GAGTCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCG 897
Db 799 ATCAGAACGCGATGTGGAATCTGATGCGCTTATTTTCGACCCGCGCAATCAAGTCG 858
QY 898 AACAGCTAGAGTGTTCAGTGGCAACAGGAGATCAATACCATTCATCCCTTACTCTCA 957
Db 859 AACCGCGAGAGTGTAGCTGTCTCTCATGTGAACCAATTACCACTTTATCCCTTATGAACA 918
QY 958 AATGTCTGAATTTGGAAGAACGAATCGCTGATTTATTTCCCTTCTGTTATCGTTCAACCA 1017
Db 919 AATGTCTGAATTTGGAAGAACGAATCGCTGATTTATTTCCCTTCTGTTATCGTTCAACCA 978
QY 1018 TTGGGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db 979 TTGGGTACAGATTCAGAACCAAGCAACCAAGTCCACCAATCGACTCCGGAACCTAGTCC 1038
QY 1078 AGSCCGCACTGACCAACCAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
Db 1039 AAGTCGCAACCTGCAACCAATCTTCAACGACTCCAGCAATCCAATTTGATGAGAAAT 1098
QY 1126 GGTAGTCACTGGTACGAAAGTTCGGGAAGGATATGTTATTCGAGAAAGGGCATCTC 1185
Db 1099 GGTCAAGAGCTGTTGGAAGAGTAGGCGATGTTATGCTTTGAGGAGAAATGGAGTTTC 1158
QY 1186 TCGTTATGCTTTGGGAAGATTTACCATCTGAACCTGTTAAATACTTTGAAAGCAAGTT 1245
Db 1159 TCGTTATATCCCAAGCAAGATCTTTACGACAGAAACAGCAGCAGGCAATTTGATGCAAACT 1218
QY 1246 ATCAAAACAGAGAGTGTTCACACCTTAACTGCTAAAAAGAAATTTGTTCTCTCG 1305
Db 1219 GGCCAGCAGGAAGTTTATCTCATAGCTAGGACTAAGAAACCTGACCTCCCATCTAG 1278
QY 1306 TGACCAAGAAATTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
Db 1279 TGATCGAGATTTTACATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1338
QY 1366 TGNAAATAGGTCGTAATTTCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGA 1425
Db 1339 TGATAATAAGGTCGACAGGTTGATTTGAGGCTTTGGATACCTTGTGGAAAGCACTCAA 1398
QY 1426 TGATCAATCGACTAATAAGAAATAATTTGGTAGATCAATTTATGGCAATCTCTAGCAACCAAT 1485

Db 1399 GGATGTCNCAAGTGAATAAGTCAAGTTAGTGGANGATATCTTGTCTTCTTAGCTCCGAT 1458
QY 1486 TACCCATCCAGAGGACTTGGCAACCAAAATTTCTCAAATTTGAGTATACTTGAAGACGAAGT 1545
Db 1459 TCGTCATCCAGAACGTTTAGGAAACCAATATCGGCAAAATTTACCTACACTGATGATGAGAT 1518
QY 1546 TCGTATGCTCAATTTAGCTGATAGTATACAACTGTCAGATGGTTACATTTTTTGTATGAACA 1605
Db 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACAGAACGCTTATATCTTTGTATCTCG 1578
QY 1606 TGATATATCAGTATGATGAGGAGATGCATATGTAACGCTCATATGGGCCATATGTCATCTG 1665
Db 1579 TGATATACCACTGATGAGGGGATGCTTATGTAACCTCAATATGAGCCCATAGCCACTG 1638
QY 1666 GATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCACTCAAGCTCAAGCTTACTATAAGA 1725
Db 1639 GATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGCAGCCAGGCTTATGCTAAAGA 1698
QY 1726 AAAAGGTATCTTACCTCATCTCCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAG 1785
Db 1699 GAAAGGTTTGACCCCTCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 1758
QY 1786 TGCAGCAGCTATTTTACCAATCGTGTGAAGGGGAAAAAGAAATTCCTACTGTTGCACTTCC 1845
Db 1759 AGCAGAAGCTTATCTCAACCGCGTGAAGGAGCTTAAGAGGTGCGCACTTGTATGCTATGCC 1818
QY 1846 ATATATGTTTGAAGATACAGTTGAGGTTTAAAGCGGTAAATTTGATTTATCTCTCATAGGA 1905
Db 1819 TTCAATCTTCAATATACTCTAGAGTCAAAACGGTGTAGTTTATCATCACTCATTTATGA 1878
QY 1906 TCATTACCAATAATTTAAATTTGCTTGGTTTGTATGATCAACATACAAAGCTCCAAATGG 1965
Db 1879 CCATTACCAATAACATCAAAATTTGAGTGGTTTGAAGGCTTTATGAGCACTTAAAGG 1938
QY 1966 CTATACCTTGAAGATTTGTTTGGGACGATTTAGTACTAGTGAACACCTGACGAAAG 2025
Db 1939 GTATCTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAG 1998
QY 2026 TCACATCTTAATGATGATGGGCAATCCAGTGAAGCATGTTAGTGGCAAGAAAGACCA 2085
Db 1999 TCGCATTCAGATAATAGTTTGGTAACTGACGACCATGTTCAAGAAACAAAAATGG 2058
QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAGGGGATGAAGCCAGTGAAGAAACACCTG 2145
Db 2059 TCAAGCTGATACCAATCAAAACGGAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118
QY 2146 TGAGCCAGAAAGTCCCTCAAGTAGAGA 2171
Db 2119 TGAGGAGAAACCCCTCGAGAGAGA 2144

RESULT 16

US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA

ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Query Match 41.1%; Score 980.8; DB 10; Length 2639;
Best Local Similarity 67.8%; Pred. No. 2.6e-227;
Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;

QY 1 TTCTTACGAGTTGGAGCTGATCAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC 57
DB 173 TTCTTATGAACTTGGTGGTCAACCAAGCTGGTCAAGTTAAGAAAGAGTCTAATCAGATTTC 232
QY 58 CTATATAGATGGAACAAAGCGCGCAACAAAGCGAATTTGACTCTGATGAGGTTAG 117
DB 233 TTATATAGATGGTATCAGGCTGGTCAAGAGCGAAGAAATTTGACACAGATGAAGTCAG 292
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGT 177
DB 293 TAAGAGAGGGATCAACGCGCAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT 352
QY 178 CACTTCATGCGGACCACTATCATTTATCAATGGTGAAGGTTCTTATGACGCTATCAT 237
DB 353 GACCTCTCATGGAGCAATATCATTTACTATAATGGCAAGGTTCTTATGATGCCAUCAT 412
QY 238 CAGTGAAGAAATTAATCATGAAGATCCCAAACTATAAGCTTAAAGATGAGGATATGTTAA 297
DB 413 CAGTGAAGAACTTCTCATGAAGATCCCAATTAATCAGTTGAAGGATTACAGACATGTCAA 472
QY 298 TGAGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGA 357
DB 473 TGAATCAAGGGTGGTATGTTATGATTAAAGTAGACGGAATACTATGTTTACCTTAAAGA 532
QY 358 TGCTGCCACGGGTAACGTCCTGACAAAGAGGAATCAATGACAAAACAAAGAGCA 417
DB 533 TGCGCCCATCGGCAATATTCGGAACAAAGAGATTAACGTCAGAGCAGGAACA 592
QY 418 TAGTCAACATCTGGAAGTGGAACTCCAAAGAACGATGGTGTGTTGCTTGGCACGTTTC 477
DB 593 CAGTCATATCAT- - - - -AACTCAAGCAGATAATGCTGTGTCGACGACAGC 643
QY 478 GCAAGCAGCTATACAGATGATGGTTATATCTTTATGTTCTGATATCATAGAGA 537
DB 644 CCAAGCAGCTTATACAAAGGATGATGGTATATCTTCAATGATCTGATATCATTTAGGA 703
QY 538 TACTGTGATGTTATATGTTCTTCATGAGATCATTTACCAATTAATTTCTTAAGATGA 597
DB 704 CACGGGTGATGTTATGTTCTTCACGGCAGCAATTAACATTAATTTCTTAAGATGA 763
QY 598 GTTATCAGCTAGCGAGTTGGTGTGTCAGAAAGCTTCTTATCTGCTCGAGGAATCTGTC 657
DB 764 GTTATCAGCTAGCGAGTTAGTGTCTGCAAGGCT- - - - - 798
QY 658 AAATTCAGAACTATCGCGCAAAATAGCGATTAACATTCAGAAACAACTGGGTACC 717
DB 799 - - - - -ATTGGAATGGGAAGCAGGGATCTGCTCCTCTTCAAGTTCTAGTTATATATGC 850
QY 718 TTCTGTAGCAATCCAGGAATACAAATTAACAAAGCAACAAAGCAACAACTAACAG 777
DB 851 AAATCCAGTTCAACCAAGATTTGTCAGAGAACCAATCTGACTGTCACTCCCACTTATCA 910
QY 778 TCAAGCAAGTCAAGTAATGACATTTGATGATGTCCTTTGAAACAGCTCTCAAACTGCTTT 837
DB 911 TCA- - -AAATCAAGGGGAAAACATTTCAAGCCCTTTACGTTGAATTTGATGCTAAAGCCCT 967
QY 838 GAGTCAACGATGATGATCTGATGCGCTTGTCTTTGATCCAGCAACAACTCAAGTCG 897
DB 968 ATCAGAACGCGATGATGATCTGATGCGCTTATTTTCGACCCCGCAATCAAGTCG 1027
QY 898 AACAGCTAGAGGTTTTCAGTGGCCACAGCGAGATCAATTAACCTTCACTCCCTTACTCTCA 957
DB 1028 AACCGCAGAGGTTGATGCTGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 1087
QY 958 AATGCTGAAATGGAAGAACGAATTCGCTCGTATATATTCCTGCTGTTATCGTTCAACCA 1017

DB 1088 AATGCTGTAATTCGAAACAAAGAAATTCCTGCTATATATTCCTTCGTTATCGTTCAACCA 1147
QY 1018 TTGGGTACCAAGATTCAAGGCCAGAACCAACCAAGTCCACAAACCACTCCGGAACCTAGTCC 1077
DB 1148 TTGGGTACCAAGATTCAAGGCCAGAACCAACCAAGTCCACAACTGACTCCGGAACCTAGTCC 1207
QY 1078 AGCCCGCAACCTGCGACCAAAATCTTAAATAGACTCAAAATCTTC- - - - -TTT 1125
DB 1208 AAGTCTCAACCTGCGACCAAAATCTTCAACCAAGTCCAAAGCAATCCAAATGATGAGAAAT 1267
QY 1126 GGTAGTACGCTGGTACGAAAGTTGGGAAGGATATGTTATTCGAAAGAAAGGCACTCTC 1185
DB 1268 GGTCAAAAGAGCTGTTGCAAAAGATGAGCGGATGTTATGTTCTTTGAGAGGATGAGTTTC 1327
QY 1186 TCGTTATGCTCTTTGCGAAAGATTTTACCATCTGAAACTGTTTAAAAATCTTGAACCAAGTT 1245
DB 1328 TCGTTATATCCAGCCAGGATCTTTTCAGCAGAAACAGCAGCAGGATTTGATGACAACT 1387
QY 1246 ATCAAAAACAGAGAGTGTGTTTTCACACACTTTTAACTGCTTAACTGAGGCTCATAAAGCTTGT 1305
DB 1388 GGCCAAGCAGGAAAGTTTATCTCATTAAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG 1447
QY 1306 TGACCAAGAAATTTTATGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGT 1365
DB 1448 TGATCGAGAAATTTTCAATAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACT 1507
QY 1366 TGNAAATAAGGGTCTGTAATCTGATTTTCAAGCCTTAGACAAAATTTATAGAACGTTGAA 1425
DB 1508 TGATAATAAGGTCGACAAAGTTGATTTTGAAGTTTGGATTAACCTGTTGGAACTCAA 1567
QY 1426 TGATGATCGAATTAATAAGAAATTTGTTAGATGATTTATTTGGCATTTCTTAGCAACCAAT 1485
DB 1568 GGTATGCTCAAGTGAATAAGTCAAGTTAGTGGATGATATTTCTTGCTTTCTTAGCTCCGAT 1627
QY 1486 TACCATCCAGAGGACTTGGCAACCAATTTCTCAAAATTCGATGATGATGATGATGATGATGAT 1545
DB 1628 TCGTCAATCCAGAACTTTAGGAAACCAATTCGCGCAATTAACCTACACTGATGATGAT 1687
QY 1546 TCGTATTGCTCAATTTAGCTGATAAGTATACAACTGATGATGATGATGATGATGATGATGAT 1605
DB 1688 TCAAGTAGCAGTTGGCAGGCAAGTACACAAAGAGACGGTTATATCTTTGATCCTCG 1747
QY 1606 TGATATATCAGTATGAGGAGATGATATGTTAACTGATGATGATGATGATGATGATGATGATGAT 1665
DB 1748 TGATATATCCAGTATGAGGGGATGCTTATGTAATCTCAATATGACCCATAGCCACTG 1807
QY 1666 GATTGGAAGATAGCTTCTGATAAGGAAAGTTGAGCTCAAGCTCATATAAGAA 1725
DB 1808 GATTAAAGAAAGATGTTTCTGTAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAAGA 1867
QY 1726 AAAAGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAG 1785
DB 1868 GAAAGTTTGGACCTCTCTTCGACAGACCCAGGATTCAGGAAATCTAGGCAAAAGG 1927
QY 1786 TCGACGAGCTATTTACAACTCGTGAAGGGGAAACCAATTCACCTCGTTCGACTTC 1845
DB 1928 AGCAGAGCTATCTACACCGGTGAAGAGCAGCTAAGAGGTGCCACTTGTATGATGCC 1987
QY 1846 ATATATGTTGAGCATACAGTTGAGGTTAAACCGGTAATTTGATTTATTTCTCATAGGA 1905
DB 1988 TTATCATCTTCAATATCTGTAAGAGTCAAAACCGGTAGTTTAACTCATACCTCATTTATGA 2047
QY 1906 TCATTAACCATATATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 1965
DB 2048 CCATTAACCAATCAATCAAAATTTGAGTGGTTTGAAGGAGGCTTTTATGAGGACCTTAGGG 2107
QY 1966 CTATACCTTTGGAGATTTGTTTCGACAGATTAAGTACTACGTAGAAACACCTGACGAACG 2025
DB 2108 GTATAGTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG 2167
QY 2026 TCCACATCTTCAATGATGATGAGGCAATGCCAGTGAAGATGTTGTTAGGCAAGAAAGCA 2085
DB 2168 TCCGCATTCAGATAATGTTTGGTAAACGCTAGTGACCACTGTTTCGTAATAAAGGACGA 2227

QY 2086 CAGTGAAGATCCAATAAGAACTTCAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGC 2145
 Db 2228 CCAAGATAGTAACTGATGAGATAGGAACATGATGAAGTAAAGTGAAGCCAACTCACCC 2287
 QY 2146 TGAGCCAGAAG 2156
 Db 2288 TGAATCTGATG 2298

RESULT 17
 US-10-158-844-243
 ; Sequence 243, Application US/10158844
 ; Publication No. US20040029118A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kunsch et al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-R
 ; COMPUTER: Dell Latitude Pentium 3
 ; OPERATING SYSTEM: Windows 98
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/158,844
 ; FILING DATE: 03-Jun-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/961,527
 ; FILING DATE: 1997-10-30
 ; APPLICATION NUMBER: US 60/029,960
 ; FILING DATE: 1996-10-31
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hyman, Mark J.
 ; REGISTRATION NUMBER: 46,789
 ; REFERENCE/DOCKET NUMBER: PB340P1D1
 ; INFORMATION FOR SEQ ID NO: 243:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2359 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 243:
 ; US-10-158-844-243

Query Match 27.4%; Score 653.8; DB 13; Length 2359;
 Best Local Similarity 68.3%; Pred. No. 6.6e-148;
 Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

QY 1 TTCTTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
 Db 939 TTCTCTAGCTTGGAGCTTACCAAGCTGTGAGGATGAAGAGTCTTAATCGAGTTGC 998
 QY 58 CTATATAGATGGAACAAACAGCGACGCAAAAACGGAGAAATTTGACTCTCTGATGAGGTTAG 117
 Db 999 TTATATAGATGTGTATGATGAGCTGGTCAAAAGGAGAGAACTTGACACCAAGATGAAGTCAG 1058
 QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGT 177
 Db 1059 TAAAGAGGGAGGGGATCAACCGCGGACAAATTTGTTATCAAGATTAACGATCAAGGTTATGT 1118
 QY 178 CACTTTCACATGGCGACCACTATCATATTTACATGTTAGGTTCTCTATGAGCTATCAT 237
 Db 1119 GACCTTCTATGAGAGCACTTATCATTAATGCGAAGGTTCTTATGATGCCATCAT 1178
 QY 238 CAGTGAAGAAATTTACTCATGAAGATCCAACTATAAGCTAAAAGATGAGGATATTGTTAA 297

Db 1179 CAGTGAAGAGCTCTCTCATGAAGATCCGAATTTATCAGTTGAAGATTCAGACATTTGTCAA 1238
 QY 298 TGAGGTCAAGGGTGCATATGTTATCAAGGTAGATGGAATAACTACTATGTTTACCTTAAGGA 357
 Db 1239 TGAATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA 1298
 QY 358 TGCTGCCACGCGGATAACGCTCCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCA 417
 Db 1299 TGCAGCTCATGCGGATAATATTCGGACAAAGAGAGATTAAACGTCAGAGAGGAGACG 1358
 QY 418 TAGTCAACATCGTGAAGGTGAACCTCAAGAAACGATGGTGTCTGTTGCTTGGCACGTC 477
 Db 1359 CAGTCATAATAAAT-----CAAGAGCAGATAAATGCTGTTGTCGAGCCAGAGC 1409
 QY 478 GCAAGGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTCATATCATAGAGGA 537
 Db 1410 CCAAGGAGGTTATACAACGGATGATGGTATATCTTCAATGATCTGATCATATGAGGA 1469
 QY 538 TACTGGTGTGCTTATATCTGCTCATGAGATCATTACCATTTACATTTCTTAAGAATGA 597
 Db 1470 CACGGGTGATGCTTATATCTGCTCACGGCAGCATTTACCATTTACATTTCTTAAGAATGA 1529
 QY 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTGCTGCGAGAAATCTGTC 657
 Db 1530 GTTATCAGCTAGCGAGTTAGCTGCTGCAAGAGCTATTGGAATGG----- 1574
 QY 658 AAATTCAAGAACTATCGCGACAAAATAGCGATAACACTTTCAAGAACAACTGGGTACC 717
 Db 1575 -GAAGCAGGATCTCGTCTCTTCAAGTCTTAGTTATATGCAATCCAGCTCAACCAA 1633
 QY 718 TTCTGTAAAGCAATCCAGGAATCAAAATACTAAACAAGCAACAACAGCAAACTAACAG 777
 Db 1634 GATTGTCAAGAACTCAATCTGACTGT-----CACTCCAACTTA 1673
 QY 778 TCAAGCAAGTCAAGTAATGACATTCATAGTCTCTTGAACAGCTCTACAACCTGCTTT 837
 Db 1674 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATGTTGATGCTTAAACCTT 1733
 QY 838 GAGTCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCACAATCAACAGTCG 897
 Db 1734 ATCAGAACCCATGTGGAATCTGATGGCTTATTTGACCCAGCGCAATCAACAGTCG 1793
 QY 898 AACAGCTAGAGGTGTGTCAGTGCACACGAGATCATTACCACTTCCATCCCTTACTCTCA 957
 Db 1794 AACCCGACAGGTGTAGCTGTCCCTCATGGTAACCATTTACCACCTTTATCCCTTATGAACA 1853
 QY 958 AATGCTGTAATTTGAAGAGAGCAATCGCTGATATTTCCCTTCTGTTATCGTTTCAACCA 1017
 Db 1854 AATGCTGTAATTTGAAGAAAACGAATGCTGATATTTCCCTTCTGTTATCGTTTCAACCA 1913
 QY 1018 TTGGGTACCAGATTCAGGGCCAGAAACAAAGTCCACAAACCGACTCCCGGAACCTAGTCC 1077
 Db 1914 TTGGGTACCAGATTCAGAGCAGAGAGAACCAAGTCCACAAACCGACTCCAGAACCTAGTCC 1973
 QY 1078 AGGCGCGCACTGACCAAAATCTTAAATAGACTCAAAATCTTTCTTGGTTAGTCAGCT 1137
 Db 1974 AAGTCCGCAACACAGCTCCAAAGCAATCCAATTTGATGAGAA-----ATTGGTCAAGAAGC 2027
 QY 1138 GGTACGAAAAGTTGGGAGGATATGTTATCGAAGAAAGGCGATCTCTCGTTATGTCCT 1197
 Db 2028 TGTTCGAAAAGTAGGCGATGGTTATGTTCTTTGAGGAGATGGAGTTTCTCGTTATATCC 2087
 QY 1198 TCGGAAAGATTTACCATCTGAACTGTTAAATACTCTGAAGCAAGTATTCAAAAACAGA 1257
 Db 2088 AGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACTGGCCACAGAGA 2147
 QY 1258 GAGTGTTCACACACTTTTAACTGCTTAAAGAAAAGAAATGTTGCTCTCGTGCACCAAGATT 1317
 Db 2148 AAGTTTATCTCATAGCTAGGAACCTTAAGAAAACAGTACCTCCCATCTAGTATCGAGATT 2207
 QY 1318 TTATGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATATAAGGG 1377

Db 2208 TTACATAAGCGTTATGACTTACTAGCAAGATTCACCAAGATTTACTTGCATATAAAGG 2267
QY 1378 TCCTAAATCTGATTTCCAGCCTTAGACAAATATTATGAGACGTTGATGATGATGAC 1437
Db 2268 TCAGCAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAAACGACTCAAGGATGTCTCAAG 2327
QY 1438 TAATAAGAAATTTGGTAGATGATTTTATGG 1469
Db 2328 TGATAAAGTCAAGTTAGTGGAAAGATATTCTTG 2359

RESULT 18
US-09-765-272-181
; Sequence 181, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-765-272-181

Query Match 16.1%; Score 385.4; DB 9; Length 1342;
Best Local Similarity 65.3%; Pred. No. 7.5e-83;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAAATAATCGTGTTCCTATATAGATGGAACCAAGCGACGCAAAAACCGAGAA 96
Db 28 TAAGGACAAATAATCGTGTCTCTATGTGGATGGCAGCGAGTCAAGTCAAGAAAGTGAAA 87
QY 97 TTTGACTCTGTAGAGTTAGCAAGCGTGAAGAAATCAATGCTGACCAATCGTCATCAA 156
Db 88 CTTGACACCAAGCAGGTTAGCCAGAAAGAAATTCAGGCTGACCAATTTGTAATCAA 147
QY 157 GATAACAGCAAGGTTATGCTACTTCCATGCGGACCACTATCAATTTATCAATGGTAA 216
Db 148 AATTACAGATCAGGCTATGTAAGTCACACGCGTACCACTATCAATTAATGGGAA 207
QY 217 GGTTCCTTATGACGCTATCATGATGAGAAATTTACTCATGAAAGATCCAACTAATAGCT 276
Db 208 AGTTCTCTTATGATGCCCTCTTTAGTGAAGAACTCTTTGATGAGGATCCAAACTCAACT 267

QY 277 AAAAGATGAGGATATTTGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
Db 268 TAAAGACGCTGATTTGTTCAATGAAGTCAAGGGTGGTTATATCATCAAGGTGATGAAA 327
QY 337 ATACTATGTTTACCTTAAAGATGCTGCCACGCGGATAAGCTCGGTACAAAGAGGAAT 396
Db 328 ATATTATGCTACCTGAAAGATGAGCTCATGCTGATAATGTTTGGAACTAAAGATGAAT 387
QY 397 CAATCGACAAAACAAGAGCATATGCAACATCGTGAAGGTGGAACTCCAAGAAACGATGG 456
Db 388 CAATCGTCAAAAACAAGAACATGTCAAGATATAGAGAGTTA-----ACTC 435
QY 457 TGCTGTTGCTTGGCAGCTTCGCAAGGACGCTTACTACAGATGATGGTTATATCTTTAA 516
Db 436 TAAATGTTGCTGTAGCAAGGCTCTCAGGACGATATACGACAAATGATGTTATGCTTTAA 495
QY 517 TGCTTCTGATATCATAGAGATCTGTTGATGCTTATATGTTCTCATGAGATCAATTA 576
Db 496 TCCAGCTGATATTATCGAAGATACGGGTAATGCTTATGCTTCTCATGAGGTCACTA 555
QY 577 CCATTACATTCCCTAAGAAATGAGTTATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCT 636
Db 556 TCCTATCATTTCCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCT 615
QY 637 ATCTGCTCGAGGAAATCTGTCAAATTCAGAACTTCAAGAACTTGCAGCAAAATAGCGATACAC 696
Db 616 GGCTGGAAAAAATATGCAACCGAGTCAAGTTAAGCTATTCTTCAACAGCTAGTGACAA--- 672
QY 697 TTCAAGAACAAACTGGGTACCTTCTGTGAAGCAATCCAGGAATCAATACTAACACAAG 756
Db 673 -----TAAACGCAATCTGT 687
QY 757 CAACAGACAGCAACTAAACAGTCAAGCAAGTCAAGTAATGACATTTGATGTTCTTTGAA 816
Db 688 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAATCTCCAGAGTCTTTTGAA 747
QY 817 ACAGCTCTACAAACTGCTTTGAGTCAACGACATGTAGAAATCTGATGCGCTTGTCTTTGA 876
Db 748 GGAACCTCTATGATTCACTAGCGCCCAAGTTACAGTGAATCAGATGCGCTGCTTTTGA 807
QY 877 TCCAGCAAAATCAAGTCCAGCAAGCTAGAGGTGTTTCAGTGCACACGCGAGATCAATTA 936
Db 808 CCCTGCTAAGATTATCAGTCTGACACCAATAGGATTCGATTCGCGATGGCGACCAATTA 867
QY 937 CCACCTTCTCCTTACTCTCAAAATGCTGAATTTGGAAGAACGAATCGCTCGTATTTATCC 996
Db 868 CCACCTTATCTTACAGCAAGCTTCTGCTTGAAGAAAAGATTGCCAGAAATGGTGGC 927
QY 997 CCTTCGT 1003
Db 928 TATCAGT 934

RESULT 19
US-09-769-787-246
; Sequence 246, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 246

```

; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-246

Query Match      16.1%; Score 385.4; DB 10; Length 1455;
Best Local Similarity 65.3%; Pred. No. 7.9e-83;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCGAAGGAAACGGAGAA 96
Db 99 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCGAAGGAAACGGAGAA 158
QY 97 TTTGACTCTCTGATGAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 156
Db 159 CTTGACACCAAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 218
QY 157 GATACAGACCAAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 216
Db 219 AATTACAGATCAGGCTATGTAAGCTGACACGCTGACCACTATCAATTAATGGAA 278
QY 217 GGTTCCTTATGACGCTATCATGAGGAAATCAATGCTGAGCAAAATCGTCATCAA 276
Db 279 AGTTCTTATGATCCCTCTTATGAGGAAATCAATGCTGAGCAAAATCGTCATCAA 338
QY 277 AAAAGATGAGGATATGTTTAAAGGTTCAAGGTTGATATGTTATCAAGGTAGATGAAA 336
Db 339 TAAAGACGCTGATATGTAAGTCAAGTCAAGGTTGTTATATCATCAAGGTCGATGAAA 398
QY 337 ATACTATGTTTACCTTAAGGATGCTGCCACGCGATTAACGTCGTAACAAAGAGGAAAT 396
Db 399 ATATTATGTTCTACCTGAAGATGAGCTGATGTAATGTTGCAACTTAAAGATGAAT 458
QY 397 CAATCGACAAAACAAAGACATAGTCAACATCGTAAGGTTGGAAGTCCAGAAACGATGG 456
Db 459 CAATCGTCAAAAACAAAGACATGTCAAAGATATGAGAAGTTA-----ACTC 506
QY 457 TGTCTGTTCCTTGGCAGCTTCGCAAGGACGCTATCTACAGATGATGGTTATATCTTTAA 516
Db 507 TAATGTTGCTGTAGCAAGGTTCTCAGGACGATATACGCAAAATGATGGTTATGTTTAA 566
QY 517 TGTCTTCTGATATCATGAGGATCTGTTGATGTTTATCATGTTCTCTCATGGAGATCATTA 576
Db 567 TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGTTCTCTCATGGAGATCATTA 626
QY 577 CCATTACATTCCTTAAGATGAGTTATCATGAGGATGTTGCTGCTGCTGCAAGACCTTCCT 636
Db 627 TCACTACATTCCTTAAGATGAGTTATCTGCTAGTGAATTAGCAGGAGCTTAAAGCATCT 686
QY 637 ATCTGCTGAGGAAATCTGTCAAATTCAGAACTTACGCGACAAATAGCGATACAC 696
Db 687 GGTGGAAGAAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA--- 743
QY 697 TTCAAGAACAACTGGGTACCTTCTGTAAGCAATCCAGGAATCTAACACAAG 756
Db 744 -----TAAACAGCAATCTGT 758
QY 757 CAACACAGCAACACTAAGCTAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 816
Db 759 AGCAAAAGATCACTAGCAAGCCCAATTAATCTGAAATCTCCAGAGTCTTTTGA 818
QY 817 ACAGCTCTCAAACTGCTTTGAGTCAAGCAATGTAAGTCTGATGGCTGCTGCTTGA 876
Db 819 GGAACCTTATGATTCACCTAGGCCCAAGCTTACAGTGAATCAGATGGCTGCTGCTTGA 878
QY 877 TCCAGCAAAATCAAGTCAAGCTAGAGGTTGCAAGTGGCCACACGAGATCATTA 936
Db 879 CCCTGCTAAGATTTATCAGTCTGACCAAAATGAGTTGCAATTCGCAATGGGACCATTA 938
QY 937 CCACCTTCACTCCCTTACTCTCAATGTTCAATTTGGNAGAACGAATCGCTGTTATTTCC 996
Db 939 CCACCTTATCTTACGCAAGCTTTCTGCTCTAGGAAAGAAAGATTGCCAGAAATGGTGCC 998

```

```

QY 997 CTTTCGT 1003
Db 999 TATCAGT 1005

RESULT 20
US-09-769-744A-23
; Sequence 23, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; FILE OF INVENTION: Proteins
; TITLE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-23

Query Match      16.1%; Score 385.4; DB 10; Length 1455;
Best Local Similarity 65.3%; Pred. No. 7.9e-83;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCGAAGGAAACGGAGAA 96
Db 99 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCGAAGGAAACGGAGAA 158
QY 97 TTTGACTCTCTGATGAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 156
Db 159 CTTGACACCAAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 218
QY 157 GATACAGACCAAGGTTAGCAAGCGTGAAGGAAATCAATGCTGAGCAAAATCGTCATCAA 216
Db 219 AATTACAGATCAGGCTATGTAAGCTGACACGCTGACCACTATCAATTAATGGAA 278
QY 217 GGTTCCTTATGACGCTATCATGAGGAAATCAATGCTGAGCAAAATCGTCATCAA 276
Db 279 AGTTCTTATGATCCCTCTTATGAGGAAATCAATGCTGAGCAAAATCGTCATCAA 338
QY 277 AAAAGATGAGGATATGTTTAAAGGTTCAAGGTTGATATGTTATCAAGGTAGATGAAA 336
Db 339 TAAAGACGCTGATATGTAAGTCAAGTCAAGGTTGTTATATCATCAAGGTCGATGAAA 398
QY 337 ATACTATGTTTACCTTAAGGATGCTGCCACGCGATTAACGTCGTAACAAAGAGGAAAT 396
Db 399 ATATTATGTTCTACCTGAAGATACGGTAAATGCTTATATGTTCTCTCATGGAGATCATTA 458
QY 397 CAATCGACAAAACAAAGACATAGTCAACATCGTAAGGTTGGAAGTCCAGAAACGATGG 456
Db 459 CAATCGTCAAAAACAAAGACATGTCAAAGATATGAGAAGTTA-----ACTC 506
QY 457 TGTCTGTTCCTTGGCAGCTTCGCAAGGACGCTATCTACAGATGATGGTTATATCTTTAA 516
Db 507 TAATGTTGCTGTAGCAAGGTTCTCAGGACGATATACGCAAAATGATGGTTATGTTTAA 566
QY 517 TGTCTTCTGATATCATGAGGATCTGTTGATGTTTATCATGTTCTCTCATGGAGATCATTA 576
Db 567 TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGTTCTCTCATGGAGATCATTA 626
QY 577 CCATTACATTCCTTAAGATGAGTTATCATGAGGATGTTGCTGCTGCTGCAAGACCTTCCT 636

```

Db 627 TCACTACATTCCTCCAAAGCGATTTATCTGCTAGTGAATAGCAGCAGTAAAGCACATCT 686
 Qy 637 ATCTGGTCGAGGAATCTGCAAAATCAAGAACTTATCGCGGCAAAATAGCGATAACAC 696
 Db 687 GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAA--- 743
 Qy 697 TTCAGAACAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATCAACACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 Qy 757 CAACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATGATAGTCTCTTGAA 816
 Db 759 AGCAAGAGGATCACTAGCAGCCAGCAAAATGAAATCTCCAGAGCTCTTTGAA 818
 Qy 817 ACAGCTTACAAAATGCTTTGAGTCAACGACATGTAGAAATCTGATGCGCTTGTCTTGA 876
 Db 819 GGAATCTATGATTCACCTAGCGCCCAACGTTACAGTGAATCAGATGGCTTGTCTTGA 878
 Qy 877 TCCAGCAAAATCAACAGTCAAGCAAGTCAAGTGTTCAGTGGCCACAGGAGATCAATTA 936
 Db 879 CCCTGCTAAGATTTACGTCGTACACCAATGGAGTTCCGATTCGATGGCGACCAATTA 938
 Qy 937 CCAGCTTACCTTACTCTCAAAATGCTGAAATGGAGAACGAATCGCTCGTATTATTC 996
 Db 939 CCAGTTTATCTTACAGCAAGCTTTCGCTTAGAAGAAAGATTGCCAGAAATGGTGCC 998
 Qy 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 21
 US-10-412-862-7
 ; Sequence 7, Application US/10412862
 ; Publication No. US20040052781A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-685
 ; CURRENT APPLICATION NUMBER: US/10/412,862
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patent In Ver. 3.0
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-862-7

Query Match 16.1%; Score 385.4; DB 13; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 7.9e-83;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 Qy 37 TAAGGAAAAATATCGTGTCTTCTATATAGATGAAACAAAGCAGCGCAAAACCGGAGAA 96
 Db 99 TAAGGACATATATCGTGTCTTCTATGTGATGCGCAGCTCAAGTCAGAAAAAGTGA 158
 Qy 97 TTTGACTCTCTGATGAGGTAGCAAGCGTGAAGAAATCAATGCTGAGCAATCTCATCA 156
 Db 159 CTTGACACCAAGCAGGTAGCCAGAAAGAAAGAAATTCAGGCTGAGCAATTTGAATCA 218
 Qy 157 GATAACAGACCAAGGCTATGCTACTTACATGCGACCATATCATTTATACAAATGGTAA 216
 Db 219 AATTACAGATCAGGCGCTATGATTAACGTCACACGCTGACCATCTATCTATATGGAA 278

Qy 217 GGTTCCTTATGACGCTATCATCTAGTGAAGATTTACTCATGAAGATCCAAACTATAAGCT 276
 Db 279 AGTTCCTTATGATGCGCTCTTTAGTGAAGAACTCTTATGATGAAGATCCAAACTATCACT 338
 Qy 277 ABAAGATGAGGATATCTGTTAATGAGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAA 336
 Db 339 TAAAGACGCTGATATGTTCAATGAGTCAAGGGTGGTATATCATCAAGGTGATGAAA 398
 Qy 337 ATACTATGTTTACCTTAAGGATGTCGCCACGCGGATAACGTCCTGTCAAAAGAGGAAT 396
 Db 399 ATATTATGTTCTACCTGAAGATGAGCTCATGCTGATAAATGTTGCAACTAAAGATGAAT 458
 Qy 397 CAATCGACAAAAACAGAGCATAGTCAACATCGTGAAGTGAAGTCAAGTCAAGAAACGATGG 456
 Db 459 CAATCGTCAAAAAACAGAACATGTCAAGATAATGAGAAGTTA-----ACTC 506
 Qy 457 TGCTGTTGCTTGCACGTTTCGCAAGGACGCTATCACTACAGATGATGTTTATATCTTTAA 516
 Db 507 TAATGTTGCTGTAGCAAGTCTCAGGACGATATACGACAAATGATGTTTATGTTTAA 566
 Qy 517 TGCTTCTGATATCATAGAGGATATGTTGATGCTTATATCGTCTCATGAGATCATTA 576
 Db 567 TCCAGCTGATATATCGAAGATACGGGTAAATGCTTATATCGTCTCATGAGGTCACTA 626
 Qy 577 CCATTACATTCCTAAGATGATTTATCAGCTAGCGAGTTGGCTGCTCAGAAAGCCTTCT 636
 Db 627 TCCTACATTTCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAGACATCT 686
 Qy 637 ATCTGCTCGAGGAATCTGTCAAAATTCAGAACTTATCGCGACAAAAATAGCGATAACAC 696
 Db 687 GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAA--- 743
 Qy 697 TTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATCACTAACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 Qy 757 CAACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGATGATGATGCTCTTGAA 816
 Db 759 AGCAAAAGGATCAACTGCAAGCCAGCAAAATAAATCTGAAATCTCCAGAGCTTTTGA 818
 Qy 817 ACAGCTTACAAAATGCTTTCAGTCAACGACATGTAGAAATCTGATGCGCTTGTCTTGA 876
 Db 819 GGAATCTTATGATTCACCTAGCGCCCAACGTTACAGTGAATCAGATGGCTGCTTTGA 878
 Qy 877 TCCAGCAAAATCAACAGTCAAGTCAAGTGAAGTGTGAGTGCACGACGAGATCAATTA 936
 Db 879 CCCTGCTAAGATTTATCAGTCTGTACCAAAATGGAGTTGCGATTCGCGATGGCGACCAATTA 938
 Qy 937 CCAGTTTATCTTACTCTCAAAATGCTGAAATGGAGAACGAATCGCTCGTATTATTC 996
 Db 939 CCAGTTTATCTTACTTACAGCAAGCTTTCGCTTAGAAGAAAGATTGCCAGAAATGGTGCC 998
 Qy 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 22
 US-10-412-850-7
 ; Sequence 7, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT APPLICATION NUMBER: US/10/412,850
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048

```

; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412,850-7

```

Query Match	16.1%;	Score 385.4;	DB 16;	Length 1455;
Best Local Similarity	65.3%;	Pred. No. 7.9e-83;		
Matches 631;	Conservative 0;	Mismatches 276;	Indels 60;	Gaps 2;
QY	37	TAAGGAAATAATCGTGTTCCTATATAGATGCGAAACAGCGGACGCGAAATAAAGCGGAA	96	
Db	99	TAAGGACAAATAATCGTGCTCTTATGTGGATGGCAGCCAGTCAAGTCAGAAAAGTGA	158	
QY	97	TTTGACTCCTGATGAGGTTAGCAAGCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAA	156	
Db	159	CTTGACACCGACACGAGTTACCCAGAAAGAGGAATTCAGGCTGAGCAAAATGTATCAA	218	
QY	157	GATAACAGACCAAGGCTATGTCACTTCACATGGCGCACCTATCATTTATACAAATGGTAA	216	
Db	219	AAATTACAGATCAGGGCTATGTAAAGTCACACGGTGACCACTATCATTTACTATATGGAA	278	
QY	217	GGTTCCCTTATGACGCTATCATCAGTGAAGAAATTAATCATGAAGATCCAACTATTAAGCT	276	
Db	279	AGTTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGATCCAACTATCAACT	338	
QY	277	AAAAGATGAGGATATGTTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	336	
Db	339	TAAAGACGCTGATATGTCGAATGAAGTCAAGGGTGGTTATATCATCAAGTCCATGGAAA	398	
QY	337	ATACTATGTTTACCTTAAAGATGTCGCCACGGGATPAACGTCCTGACAAAAGAGAAAT	396	
Db	399	ATATTATGTCTACCTGAAAGATCAGCTCATGCTGATATGTTTGAACATAAGATGAAT	458	
QY	397	CAATCGACAAAAACAAGACATAGTCAACATCGTGAAGGTGGAACTTCCAAAGAACGATGG	456	
Db	459	CAATCGTCAAAAACAGAACATGTCAAAGATAATGAAAGGTTA-----ACTC	506	
QY	457	TGCTGTGTCCCTTGCCACGTTTCGCAAGGACGCTATACTACAGATGATGGTTATATCTTTAA	516	
Db	507	TAATGTGTCTGTAGCAAGGTCCTCAGGACGATATACGACAAATGATGGTTATGTCTTTAA	566	
QY	517	TGCTCTTGATATCATAGAGGATCTGGTGATGCTTATATCGTTTCCATGGAGATCAATTA	576	
Db	567	TCCAGCTGATTAATTCGAAGATACGGGTAAATGCTTATATCGTTTCCATGGAGGTCACTA	626	
QY	577	CCATTACATTCCTTAAGAAATGAGTTATCAGCTACGAGTTGGCTGCTGCAGAGACCTTCT	636	
Db	627	TCACTACATTCCCAAAAGCGATTATCTGCTAGTGAATTAGCGACGCTAAAGACATCT	686	
QY	637	ATCTGTCGAGGAAATCTGTCTAAATTCAGAAACCTATCCGCCACAAAAATAGCGATAACAC	696	
Db	687	GGCTGCAAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAA---	743	
QY	697	TTCAAGAACAAATCGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTTAACACAAG	756	
Db	744	-----TAAACGCAATCTGT	758	
QY	757	CAACAAACAGCAACACTAAACAGTCAAGCAAGTCAAAAGTAAATGATGATGCTCTTGA	816	
Db	759	AGCAAAAGGATCAACTAGCAAGCCAGCAATTAATCTGAAATCTCCAGAGTCTTTTGA	818	
QY	817	ACAGCTCTACAACTCGCTTTGAGTCAACGACATGTAGAAATCTGATGGCTTGTCTTTGA	876	
Db	819	GGAACTCTATGATTCACCTAGCGCCCAACGTTTACAGTGAATCAGATGGCTTGCTTTGA	878	
QY	877	TCCAGCACAAATCAAAAGTCGAAACAGCTAGAGGTGTTCAGTGGCCACACGAGATCAATTA	936	
Db	879	CCCTGCTAAGATTATCAGTCGTACACCAATGGAGTTGCGATTCCGCGATGCGACCAATTA	938	

Qy	937	CACTTCATCCCTTACTCTCAATGTC	TGAAATCGAAGAACGAATCGCTCGTATTATTC	936
Db	939	CACTTTATCTCTTACAGCAAGCTTCTGCCTT	TAGAAGAAAGANTGCCAGATGGTGCC	938
Qy	997	CCTTCGT	1003	
Db	999	TATCAGT	1005	

```

RESULT 23
US-10-387-783-7
; Sequence 7, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387, 783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-7

```

Query Match	16.1%;	Score 385.4;	DB 16;	Length 1455;
Best Local Similarity	65.3%;	Pred. No. 7.9e-83;		
Matches	631;	Conservative 0;	Mismatches 276;	Indels 60; Gaps 2;
QY	37	TAAGGAAATAATCGTCTTTTCCTATATAGATGGAAAAA	CAAGCGACGCAAAAAACGGGAA	96
Db	99	TAAGGCAATAATCGTCTCTTTATGTGGATGGCAGCAGT	CAAGTCAAGAAAGTGA	158
QY	97	TTTGACTCCTCATGAGGCTTAGCAAGCGTTAGCGAAG	GAATCAATGCTCAGCAAAATCGTCATCAA	156
Db	159	CTTGACACCAAGCCAGGTTAGCCAGAAAGAGAAATTC	AGCTGAGCAAAATTTGAATCAA	218
QY	157	GATTAACAGACCAAGGCTATGCTACTTCACATGGCGAC	CACTATCATTTATTACAAATGGTAA	216
Db	219	AATTACAGATCAGGGCTATGTAACGTCAACGGTGACCA	TATCATTAATGGAA	278
QY	217	GGTTCCCTATGACGCTATCATCAGTGAAGAAATTA	CTCATGAAAGATCCAACTATAGCT	276
Db	279	AGTTCCCTATGATGCCCTCTTTAGTGAAGAACTCTT	GATGAAGATCCAACTATCAACT	338
QY	277	AAAAGATGAGGATATTGTTTAATGAGGTC	CAAGGTGGATGTTTATCAAGGTAGATGAAA	336
Db	339	TAAAGACGCTGATATTGTCATGAAGTCAAGGGTGGT	TATATCATCAAGTCCATGGAAA	398
QY	337	ATACTATGTTTACCTTTAAGGATGCTCCCAACGGGAT	TAACGTCGTACAAAAGAGAAAT	396
Db	399	ATATTATGTCTACCTCGAAAGATGCAGCTATGCTGA	TAAATGTTTCGAACATAAGATGAAAT	458
QY	397	CAATCGACAAAAACAAGAGCATAGTCAACATCGTGA	AGTGGAACTCCAGAAGCATGG	456
Db	459	CAATCGTCAAAAACAAGAACATGTCANAGATTA	ATGAAAGGTTA-----ACTC	506
QY	457	TGCTGTGTGCTTGGCACGTTTCGCAAGGACGCTATA	CTACAGATGATGGTTATATCTTTAA	516
Db	507	TAAATGTTGCTGTAGCAAGGCTCTCAGGACGATAT	ACGCAAAATGATGGTTATGTCCTTTAA	566
QY	517	TGCTCTGTATATCATAGAGGATCTGGTGATGCTTAT	ATCGTTCTCATGGAGATCAATTA	576
Db	567	TCAGCTGTATATTTGGAAGATACGGGTAAATGCTTA	TATCGTTCTCATGGAGGTCACTA	626

QY 577 CCATTACATCTTAAGAAATGAGTTATCAGCTAGCGAGTTGGCTGCTGAGAGCCTTCCT 636
 Db 627 TCCTACATTTCCCAAAAGCGATTTATCTGCTAGTGAATTTAGCAGCAGCTAAAGCACATCT 686
 QY 637 ATCTGGTCGAGGAATCTGTCAAATTCAGAACTATCGCGACAAAATACGGATAACAC 696
 Db 687 GGCTGGAAAAAATATGCAACCGAGTCAAGTTAAGCTATTCTTCAACAGCTAGTGAACA 743
 QY 697 TTCAAGAACAAATCGGTGACCTCTGTAGCAATCCAGAACTACAAATACATAACACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 QY 757 CAACAACAGCAACACTAAACAGTCAAGCAAGTCAAGAGTAAATGATGATCTCTTTGAA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAATCTCCAGAGCTCTTTTGA 818
 QY 817 ACAGCTCTACAACTGCTTTGAGTCAAGCATGTAGAACTGATGCTGCTTGTGA 876
 Db 819 GGAATCTATGATTCACCTAGGCGCAAGTTACAGTGAATCAGATGCGCTGGTCTTGA 878
 QY 877 TCCAGCAAAATCAAACTCGAAGCTAGAGGTTGTCAGTGGCCACACGGAGATCAITA 936
 Db 879 CCCTGCTAAGATTATCAGTGTGACCAAAATGGAGTTGCGATTCCGCATGGCGACCAITA 938
 QY 937 CCACTTCATCCCTTACTCTCAATGTCTGAATTTGGAAGCAAAATCGCTGATTATTCC 996
 Db 939 CCACCTTTATCTCTACAGCAAGCTTTCTGCTTTAGAAAGAAAGATTGCCAGAAATGGTGCC 998
 QY 997 CTTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 24
 US-09-884-465A-1
 ; Sequence 1, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 3120
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-884-465A-1
 Query Match 16.1%; Score 385.4; DB 10; Length 3120;
 Best Local Similarity 65.3%; Pred. No. 1.2e-82;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAGGAAATTAATCGTGTTCCTTATATAGATGGAATAAAGAGGCGGCAAAACGGAGAA 96
 Db 99 TAAGGCAATAATCGTGTCTCTTATGTTGGATGGCAGCCAGTCAAGTCAGAAAAGTGA 158
 QY 97 TTTGACTCTGATGAGCTAGCAGCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAA 156
 Db 159 CTTGACACCAAGCAGGTTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATGTAATCAA 218
 QY 157 GATAACAGACCAAGGCTATGTCACTTCACATGGCGACCATATCAATTATTAATGTGTA 216

Db 219 AATTACAGATCAGGGCTATGTAACGTCAAGGTGACCACTATCAATTACTATATGGAA 278
 QY 217 GGTTCCTTTATGACCTATCATCTAGTGAAGAAATTAATCATGAAGATCCAACTATAGCT 276
 Db 279 AGTTCTTTATGATGCCCTCTTTAGTGAAGAACTCTTGTATGAAGGATCCAACTATCACT 338
 QY 277 AAAAGATGAGGATATGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGTGA 336
 Db 339 TAAAGACGCTGATATTGTTCAATGAAGTCAAGGGTGGTATATATCATCAAGGTGCGATG 398
 QY 337 ATACTATGTTTACCTTAAGGATGCTGCCAGCGGATAACGTCCTGTAACAAAGAGGAAT 396
 Db 399 ATATTATGTTTACCTTAAGGATGCGCTCATGCTGATGATGTTTCAACTAAGATGAAT 458
 QY 397 CAATCGCAAAAACAAAGAGCATAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGG 456
 Db 459 CAATCGTCAAAAACAAAGACATGTCNAAGATAATGAGAAGTTA-----ACTC 506
 QY 457 TGCTGTTGCCCTTGGCAGCTTCGCAAGGACGCTTACTACAGATGATGCTTATCTTTAA 516
 Db 507 TAATGTTGCTCTAGCAAGGTCTCAGGAGCAGATATAACGACAAATGATGTTATGTTTAA 566
 QY 517 TGCTTCTCATATCATAGAGGATGCTGCTGATGCTTATATCGTTCCTCATGAGATCAITA 576
 Db 567 TCCAGCTGATATTTATCGAAGATACCGGTAATGCTTATATCGTTCCTCATGAGGTCATA 626
 QY 577 CCATTACATCTCTAAGAAATGATTTATCAGCTAGCGAGTGGCTGTCGAGAAAGCCTTCCT 636
 Db 627 TCCTACATTTCCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCT 686
 QY 637 ATCTGCTGAGGAAATCTGTCAATTCAGAACTTATCGCCGACAAAATAGGATAACAC 696
 Db 687 GGCTGGAATAAATATGCAACCGAGTCAAGTAAAGCTATTCTTCAACAGCTAGTGACAA 743
 QY 697 TTCAAGAACAAACTGGTACCTTCTGTAGCAATCAGGAACTCAAAATACTAACAACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 QY 757 CAACAACAGCAACACTTAACAGTCAAGCAAGTCAAGAGTAAAGTATGATGCTCTTGA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAATCTCCAGAGTCTTTTGA 818
 QY 817 ACAGCTCTAACAATGCTTTGAGTCAAGCATGATGATGATCTGATGCTGCTTGTCTTGA 876
 Db 819 GGAATCTATGATTCACCTAGGCGCAACGTTACAGTGAATCAGATGGCTGCTGCTTGA 878
 QY 877 TCCAGCAAAATCAAACTCGAAGTCAAGCTAGAGGTTGTCAGTGGCCACACGGAGATCAITA 936
 Db 879 CCCTGCTAAGATTATCAGTCCGTACACCAATGGAGTTGCGATTCGCGACCAITA 938
 QY 937 CCACTTCATCCCTTACTCTCAAAATGTCGAATTTGGAAGCAAGAAATCGCTGATTATTCC 996
 Db 939 CCACCTTTATCTCTACAGCAAGCTTTCTGCTTTAGAAAGAAAGATTGCCAGAAATGGTGCC 998
 QY 997 CTTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 25
 US-09-884-465A-2
 ; Sequence 2, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A

; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 5048
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-2

Query Match 16.1%; Score 385.4; DB 10; Length 5048;
 Best Local Similarity 65.3%; Pred. No. 1.6e-82;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY	37	TAAGGAAATATCGTGTTCCTATATAGTGGAAACAAAGCGGACGCAAAAACGGAGAA	96
DB	1875	TAAGGACATATATCGTGTCTTTATGTGGTGGCGCCAGTCAAGTCAGAAAAGTGAAA	1934
QY	97	TTTGACTCTCGATGAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAA	156
DB	1935	CTTGACACACAGCCAGGTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATGTAATCAA	1994
QY	157	GATAACAGACCAAGCTATGTCATTCACATGGCGACCACTATCATTTATTAATGGTAA	216
DB	1995	AATTACAGATCAGGCTATGTAAGCTCACACGGTGACCACTATCATTTACTATAATGGAA	2054
QY	217	GGTTCCTTATGACCTATCATCAGTGAAGAAATTAATGCTGAAGAAATCGTGAAGAAAT	276
DB	2055	AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGAAGAAATCGTGAAGAAAT	2114
QY	277	AAAGATGAGGATATGTTAATAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	336
DB	2115	TAAAGACGCTGATATGTCATGAAGTCAAGGGTGGTATATCATCAAGGTGATGAAA	2174
QY	337	ATACTATGTTTACCTTAAGGATGCTGCCAGCGGATAAGCTCCGTACAAAAGAGGAAAT	396
DB	2175	ATATATGCTCTGCTGAAGATGAGCTCATGCTGATAATGTTTGAAGTAAAGATGAAAT	2234
QY	397	CAATCGCAAAAAACAGAGCATAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGG	456
DB	2235	CAATCGTCAAAAACAGAAATGTCAAAGATATGAGAAGTTA-----ACTC	2282
QY	457	TGCTGTTCCTTGGCAGTTCGCAAGGACGCTATATACAGATGATGGTTATATCTTTAA	516
DB	2283	TAATGTTGCTGTAGCAAGGTCTCAGGGACGATATACGACAAATGATGGTTATGCTTTAA	2342
QY	517	TGCTTCTGATATCATAGAGGATAGTGGTATGCTTATATCGTTCCTCATCGAGATCATTA	576
DB	2343	TCAGCTGATATTTATCGAAGATACGGGTAAATGCTTATATCGTTCCTCATGGAGTCACTA	2402
QY	577	CCATTACATTCCTTAAGATGAGTTATCAGCTAGGAGTGGCTGCTGCAGAAAGCCTTCCT	636
DB	2403	TCACTACATTCCTCAAAAGGATTTATCTGTAGTAAATAGCAGAGCTAAAGCAATCT	2462
QY	637	ATCTGTCGAGGAAATCTGTCAAATTCAGAACTATCGCCGACAAAATAGCGATAACAC	696
DB	2463	GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA---	2519
QY	697	TTCAGAAACAACTGGGTACCTTCTGTAAGCAATCCAGGAATACAAATACTAACACAAG	756
DB	2520	-----TAACACGCAATCTGT	2534
QY	757	CAACACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAAATGATGATGATCTTTGAA	816
DB	2535	AGCAAAAGGATCAACTAGCAAGCAGCAAAATAACTGAAAATCTCCAGAGCTTTTGA	2594
QY	817	ACAGCTCTCAAACTCCCTTTGAGTCAACGACATGATGATGATGATGGCTTGTCTTTGA	876
DB	2595	GGAACCTCTATGATTCACCTAGCGCCCAACGTTACAGTGAATCAGATGGCTGCTTTGA	2654
QY	877	TCCAGCAAAATCACAGTTCGAACGCTAGAGGTGTTGAGTGGCCACACGAGATCATTA	936

Db	2655	CCCTGCTAAGATTATCAGTGTGATACCAAAATGGAGTTGGATTCGGCATGGGACCATTA	2714
QY	937	CCACTTCATCCCTTACTCTCAAAATGCTGAATTTGGAAGAACGAATCGCTCGTATTATCC	996
Db	2715	CCACTTTATTCCTTACAGCAAGCTTTCTGCTTTAGAGAAAGATTGCCAGATGGTGCC	2774
QY	997	CCTTCGT	1003
Db	2775	TATCAGT	2781

Search completed: September 30, 2004, 22:39:20
 Job time : 1157 secs

ID AAW55090 standard; protein; 796 AA.

XX AAW55090;
AC
XX 02-OCT-1998 (first entry)
DT
XX Streptococcus pneumoniae SP0036 protein.
DE
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX Streptococcus pneumoniae; otitis media; meningitis.
KW
XX Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
FT Misc-difference 456
FT /label= unknown
FT /note= "encoded by GNA"
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX WPI; 1998-272224/24.
DR N-PSDB; AAV27351.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 59-60; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
XX Sequence 796 AA;
SQ
Query Match 100.0%; Score 4163; DB 2; Length 796;
Best Local Similarity 100.0%; Pred. No. 9.4e-299;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRNSVSDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLYQARTVKNRNSVSDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISSEELMKDPNFKDIEDIVNEVKGQYVUKVGGYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISSEELMKDPNFKDIEDIVNEVKGQYVUKVGGYVYLKDA 120
QY 121 AHADNVTKEINQKQHSQHRGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
Db 121 AHADNVTKEINQKQHSQHRGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIPKNELASASELAFAFLSGRGNLSNSRTYRRQNSDNTSRTNWPS 240

Db 181 GDAYIVPHGDHYHYIPKNELASASELAFAFLSGRGNLSNSRTYRRQNSDNTSRTNWPS 240
QY 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQYLKPLSQRHVESDGLVDFPAQITSRT 300
Db 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQYLKPLSQRHVESDGLVDFPAQITSRT 300
QY 301 ARGVAVPHGDHYHYIPYSQMSSELEERARIIPLYRSNHVWVDSRPPSPQPTPEPSG 360
Db 301 ARGVAVPHGDHYHYIPYSQMSSELEERARIIPLYRSNHVWVDSRPPSPQPTPEPSG 360
QY 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLFSETVKLESKLSKQES 420
Db 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLFSETVKLESKLSKQES 420
QY 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAPLAPITHTPERLKGKNSQIEYTEDVRIAQLADKYTTSDGYIFDHDHDIISD 540
Db 481 KEKLVDDLLAPLAPITHTPERLKGKNSQIEYTEDVRIAQLADKYTTSDGYIFDHDHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKIGILPSPDADVKNPTGDSAAAIY 600
Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKIGILPSPDADVKNPTGDSAAAIY 600
QY 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYNIKEAFDDHHTYKAPNGYTTLED 660
Db 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYNIKEAFDDHHTYKAPNGYTTLED 660
QY 661 LPAITKYVVEHPDERPHSDGNGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
Db 661 LPAITKYVVEHPDERPHSDGNGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
QY 721 QVETEKVBAQLKEAEVLLAKVTDSSLKANATETLAGLRNLTQIMDNNSIMAEAEKLLA 780
Db 721 QVETEKVBAQLKEAEVLLAKVTDSSLKANATETLAGLRNLTQIMDNNSIMAEAEKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796
RESULT 2
ID ABP54584 standard; protein; 796 AA.
XX ABP54584;
AC ABP54584;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP036 protein sequence SEQ ID NO:56.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2002-479261/51.
 DR N-PSDB; ABQ84819.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX
 PS Claim 11; Page 27; 70pp; English.
 PS
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX
 SQ Sequence 796 AA;
 Query Match 100.0%; Score 4163; DB 5; Length 796;
 Best Local Similarity 100.0%; Pred. No. 9.4e-299;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60
 Db 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60
 QY 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEIVNEVKGKGVYKVDGKYVYLKDA 120
 Db 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEIVNEVKGKGVYKVDGKYVYLKDA 120
 QY 121 AHADNVTKEINRQKQHSQHREGGTPRNDGVALARSQGYTTDDGYIFNADSIIEDT 180
 Db 121 AHADNVTKEINRQKQHSQHREGGTPRNDGVALARSQGYTTDDGYIFNADSIIEDT 180
 QY 181 GDVIVPHGDHYHVPKNEISASLAAAEAPLSGRGNLSRTYRRQNSDTSRTNWPS 240
 Db 181 GDVIVPHGDHYHVPKNEISASLAAAEAPLSGRGNLSRTYRRQNSDTSRTNWPS 240
 QY 241 VSNPGTINTNTSNNTNSQASQNDIDSLLKQYKLPQSRHVSGLVDFPAQITSRT 300
 Db 241 VSNPGTINTNTSNNTNSQASQNDIDSLLKQYKLPQSRHVSGLVDFPAQITSRT 300
 QY 301 ARGVAVPHGDHYHVPYQMSSELEERARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 360
 Db 301 ARGVAVPHGDHYHVPYQMSSELEERARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 360
 QY 361 PQAPNLKIDSNSLSVQLVRKVGEGVYFEEKIGISRYVFAKOLPSETVKNLESKLKQES 420
 Db 361 PQAPNLKIDSNSLSVQLVRKVGEGVYFEEKIGISRYVFAKOLPSETVKNLESKLKQES 420
 QY 421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKNGNSDFQALDKLLERLNDESTN 480
 Db 421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKNGNSDFQALDKLLERLNDESTN 480
 QY 481 KEKLVDDLLAFLAPITHPERLGNPSQIETVEDEVRIQLADKYTTSDGYIFEHDIISD 540
 Db 481 KEKLVDDLLAFLAPITHPERLGNPSQIETVEDEVRIQLADKYTTSDGYIFEHDIISD 540
 QY 541 EGDVYVTPHMGSHWIGKDSLSKKEVAQAQYKEKGIILPPSPDADVKANPTGDSAAATY 600
 Db 541 EGDVYVTPHMGSHWIGKDSLSKKEVAQAQYKEKGIILPPSPDADVKANPTGDSAAATY 600
 QY 601 NRKVGKRIPLVRLPWVHTVEVKNGNLIIPKQHYHNIKEAFDDHTYKAPNGYTTLED 660
 Db 601 NRKVGKRIPLVRLPWVHTVEVKNGNLIIPKQHYHNIKEAFDDHTYKAPNGYTTLED 660

QY 661 LFATIKYVYEHDPDRPHSDNGWGNASEHVLGKKHSEDPNKNFKADEEPVEETPAEPEVP 720
 Db 661 LFATIKYVYEHDPDRPHSDNGWGNASEHVLGKKHSEDPNKNFKADEEPVEETPAEPEVP 720
 QY 721 QVETEKVEAQLKEAEVLLAKVTDSSILKANATETLAGLRNNLTQIIMDNNSIMAEKLLA 780
 Db 721 QVETEKVEAQLKEAEVLLAKVTDSSILKANATETLAGLRNNLTQIIMDNNSIMAEKLLA 780
 QY 781 LLKGSNPFSSVSKEKIN 796
 Db 781 LLKGSNPFSSVSKEKIN 796
 RESULT 3
 ADC45137
 ID ADC45137 standard; protein; 796 AA.
 XX
 AC ADC45137;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE S. pneumoniae antigenic protein SP036.
 XX
 KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US6573082-B1.
 PD 03-JUN-2003.
 XX
 PF 28-MAR-2000; 2000US-00536784.
 XX
 PR 31-OCT-1996; 96US-0029960P.
 PR 30-OCT-1997; 97US-00961083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX
 DR WPI; 2003-764574/72.
 DR N-PSDB; ADC45136.
 XX
 PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by Streptococcus pneumoniae.
 XX
 PS Example 1; SEQ ID NO 56; 58pp; English.
 XX
 CC The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence
 CC represents an S. pneumoniae antigenic protein.
 XX
 SQ Sequence 796 AA;
 Query Match 100.0%; Score 4163; DB 7; Length 796;
 Best Local Similarity 100.0%; Pred. No. 9.4e-299;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60
 Db 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLKDIEDIVNEVKGQYVTKDQKYYVYLKDA 120
 Db 61 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLKDIEDIVNEVKGQYVTKDQKYYVYLKDA 120
 QY 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGQRYTDDGYIFNASDIIEDT 180
 Db 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGQRYTDDGYIFNASDIIEDT 180
 QY 181 GDVIVPHGDHYHYIPKNELASASELAFAFLSGRGLNSRTRYRRQNSDNTSRNWVPS 240
 Db 181 GDVIVPHGDHYHYIPKNELASASELAFAFLSGRGLNSRTRYRRQNSDNTSRNWVPS 240
 QY 241 VSNPQTNTNTSNNSTNSQASQNDISLILKQYKPLPSORHVESDGLVFPDPAQITSR 300
 Db 241 VSNPQTNTNTSNNSTNSQASQNDISLILKQYKPLPSORHVESDGLVFPDPAQITSR 300
 QY 301 ARGVAVPHGDHYHETPYQMSSELEBRIARIIPLRYSNHWVPDSRPEQSPQTPPEPSFG 360
 Db 301 ARGVAVPHGDHYHETPYQMSSELEBRIARIIPLRYSNHWVPDSRPEQSPQTPPEPSFG 360
 QY 361 QPAPNLKIDNSSLVSQVRKVGEGYVPEEKIGSYVFAKDLPSRTVKNLESKLSQES 420
 Db 361 QPAPNLKIDNSSLVSQVRKVGEGYVPEEKIGSYVFAKDLPSRTVKNLESKLSQES 420
 QY 421 VSHTLTAKENVAPRQDFYDQYKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
 Db 421 VSHTLTAKENVAPRQDFYDQYKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
 QY 481 KEKLVDDLLAFAPITTHPERLKGPNQSIETEDVRIQLADKYTTSDGYIFDEHDIISD 540
 Db 481 KEKLVDDLLAFAPITTHPERLKGPNQSIETEDVRIQLADKYTTSDGYIFDEHDIISD 540
 QY 541 EGDVYVTPHMGSHWIKGDSLSDEKVAQAQYVTEKGIILPPSPDADVKANPTGDSAAIY 600
 Db 541 EGDVYVTPHMGSHWIKGDSLSDEKVAQAQYVTEKGIILPPSPDADVKANPTGDSAAIY 600
 QY 601 NRKVGKRIPLVPLPMVHTVVRKGNLIIPKHQHYHNIKPAWFDHTYKAPNGYTLSD 660
 Db 601 NRKVGKRIPLVPLPMVHTVVRKGNLIIPKHQHYHNIKPAWFDHTYKAPNGYTLSD 660
 QY 661 LFATIKYVHEPDERPHSNDGWNAGSEHVLGKDHSDPNKFPKADPEEVEETPAPEVP 720
 Db 661 LFATIKYVHEPDERPHSNDGWNAGSEHVLGKDHSDPNKFPKADPEEVEETPAPEVP 720
 QY 721 QVETEKVEAQLKEAEVLLAKVTSSKANATETIAGLRNMLTLQIMDNNSIMAEKLLA 780
 Db 721 QVETEKVEAQLKEAEVLLAKVTSSKANATETIAGLRNMLTLQIMDNNSIMAEKLLA 780
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 781 LLKGSNPSSVSKEKIN 796

RESULT 4
 ABU01598 standard; protein; 802 AA.
 ID ABU01598
 XX AC ABU01598;
 XX AC ABU01598;
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 DE S. pneumoniae type 4 strain protein from coding region #1174.
 XX DE Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX OS Streptococcus pneumoniae; type 4 strain.
 XX PN WO200277021-A2.
 XX XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB002163.
 XX 27-MAR-2001; 2001GB-00007658.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX06886.
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX Claim 1; SEQ ID NO 2348; 56pp; English.
 PS The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX SQ Sequence 802 AA;

Query Match 100.0%; Score 4163; DB 6; Length 802;
 Best Local Similarity 99.9%; Pred. No. 9.5e-299;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTDPVSKREGINAEQIVIKITDQGYVT 60
 Db 7 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTDPVSKREGINAEQIVIKITDQGYVT 66
 QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLKDIEDIVNEVKGQYVTKDQKYYVYLKDA 120
 Db 67 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLKDIEDIVNEVKGQYVTKDQKYYVYLKDA 126
 QY 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGQRYTDDGYIFNASDIIEDT 180
 Db 127 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGQRYTDDGYIFNASDIIEDT 186
 QY 181 GDVIVPHGDHYHYIPKNELASASELAFAFLSGRGLNSRTRYRRQNSDNTSRNWVPS 240
 Db 187 GDVIVPHGDHYHYIPKNELASASELAFAFLSGRGLNSRTRYRRQNSDNTSRNWVPS 246

QY 241 VSNPGTNTNTSNTNSQASQSNIDISLLKQYKLPISQRHVSDGLVFPQAITSRT 300
 Db 247 VSNPGTNTNTSNTNSQASQSNIDISLLKQYKLPISQRHVSDGLVFPQAITSRT 306
 QY 301 ARGVAVPHGDHYHFIPIYSOMSELEERIIIPLRVSRNHVWVDSRPEQSPQTPPEPSG 360
 Db 307 ARGVAVPHGDHYHFIPIYSOMSELEERIIIPLRVSRNHVWVDSRPEQSPQTPPEPSG 366
 QY 361 POPAPNLKIDSNSLSVQSVRRVGVGVEEKGISRYVFAKDLPSSTVKNLESLSKQES 420
 Db 367 POPAPNLKIDSNSLSVQSVRRVGVGVEEKGISRYVFAKDLPSSTVKNLESLSKQES 426
 QY 421 VSHTLTAKENVAPRQDEYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
 Db 427 VSHTLTAKENVAPRQDEYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 486
 QY 481 KEKLVDDLLAFAPITTHPERLGKPNQJOIETDEVRILAQADKYTTSQYIPDEHDIISD 540
 Db 487 KEKLVDDLLAFAPITTHPERLGKPNQJOIETDEVRILAQADKYTTSQYIPDEHDIISD 546
 QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKNPTGDSAAAIY 600
 Db 547 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKNPTGDSAAAIY 606
 QY 601 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKDHYNHNIKFAWPDHHTYKAPNGYTTLED 660
 Db 607 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKDHYNHNIKFAWPDHHTYKAPNGYTTLED 666
 QY 661 LFAITKYVVEHPDRPHNSDGNWGNASEHVLGKKHSEDPNKNFKADEPVEETPAEPEVP 720
 Db 667 LFAITKYVVEHPDRPHNSDGNWGNASEHVLGKKHSEDPNKNFKADEPVEETPAEPEVP 726
 QY 721 QVETEKVEAQKAEVLLAKVTDSSKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 780
 Db 727 QVETEKVEAQKAEVLLAKVTDSSKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 786
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 787 LLKGSNPSSVSKEKIN 802

RESULT 5
 AAB01468
 ID AAB01468 standard; protein; 819 AA.
 AC AAB01468;
 XX
 20-OCT-2000 (first entry)
 DE Recombinant variant of Sp36 (Sp36A) of *S. pneumoniae*.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 XX meningitis; lobar pneumonia.
 OS Streptococcus pneumoniae.
 XX

Key Location/Qualifiers
 FT Region 63..68
 FT /label= Histidine triad residue
 FT Region 118..145
 FT /label= Coiled coil region
 FT Region 189..194
 FT /label= Histidine triad residue
 FT Region 309..314
 FT /label= Histidine triad residue
 FT Region 406..434
 FT /label= Coiled coil region
 FT Region 462..493
 FT /label= Coiled coil region
 FT Region 550..555
 FT /label= Histidine triad residue

FT Region 634..639
 FT /label= Histidine triad residue
 FT Region 724..751
 FT /label= Coiled coil region
 PN WO200037105-A2.
 PD 29-JUN-2000.
 PF 21-DEC-1999; 99WO-US030390.
 XX 21-DEC-1998; 98US-0113048P.
 PR (MEDI-) MEDIMUNE INC.
 PA
 XX Johnson LS, Koenig S, Adamou JE;
 PI WPI; 2000-452129/39.
 DR N-PSDB; AAA47604.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX
 PS Claim 1; Page 61-64; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of *S. pneumoniae* in order to utilize such polypeptides in vaccines to
 CC protect against a wide variety of *S. pneumoniae*. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX
 SQ Sequence 819 AA;
 Query Match 100.0%; Score 4163; DB 3; Length 819;
 Best Local Similarity 99.9%; Pred. No. 9.8e-299;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTDPDEVSKREGINAEQIVIKITDQGYVT 60
 Db 21 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTDPDEVSKREGINAEQIVIKITDQGYVT 80
 QY 61 SHGDHYHYNGKVPYDAIIIEELLMKDPNPKLDEIVNEVGKGYIVKVDGKYVYLKDA 120
 Db 81 SHGDHYHYNGKVPYDAIIIEELLMKDPNPKLDEIVNEVGKGYIVKVDGKYVYLKDA 140
 QY 121 AHADNVRTKEEINRQKQHSQHREGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 Db 141 AHADNVRTKEEINRQKQHSQHREGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 200
 QY 181 GDAYIVPHGDHYHYIPKNELSAELAAEAFISGRNLSNRTYRQNSDNTSRTNWVPS 240
 Db 201 GDAYIVPHGDHYHYIPKNELSAELAAEAFISGRNLSNRTYRQNSDNTSRTNWVPS 260
 QY 241 VSNPGTNTNTSNTNSQASQSNIDISLLKQYKLPISQRHVSDGLVFPQAITSRT 300
 Db 261 VSNPGTNTNTSNTNSQASQSNIDISLLKQYKLPISQRHVSDGLVFPQAITSRT 320
 QY 301 ARGVAVPHGDHYHFIPIYSOMSELEERIIIPLRVSRNHVWVDSRPEQSPQTPPEPSG 360
 Db 321 ARGVAVPHGDHYHFIPIYSOMSELEERIIIPLRVSRNHVWVDSRPEQSPQTPPEPSG 380
 QY 361 POPAPNLKIDSNSLSVQSVRRVGVGVEEKGISRYVFAKDLPSSTVKNLESLSKQES 420
 Db 381 POPAPNLKIDSNSLSVQSVRRVGVGVEEKGISRYVFAKDLPSSTVKNLESLSKQES 440

RESULT 7	
AAB12727	
ID	AAB12727 standard; protein; 821 AA.
XX	
AC	AAB12727;
XX	
DT	21-NOV-2000 (first entry)
XX	

CC	antigen, from the present invention	
XX		
QQ	Sequence 821 AA:	
	Query Match	77.3%; Score 3218; DB 3; Length 821;
	Best Local Similarity	75.2%; Pred. No. 8.2e-229;
	Matches 616; Conservative	65; Mismatches 104; Indels 34; Gaps 3;
QY	1 SYELGLVQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT	60
DB	2 AYELGLHQAGTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT	61
QY	61 SHGDHYHYNGKVPYDAIISELLMKDPNKLKDEIVNFKGGYIVKDGKVVYVLKDA	120
DB	62 SHGDHYHYNGKVPYDAIISELLMKDPNQLKXSDIVNEIKGGYIVKNGKYYVTLKDA	121
QY	121 AHADNVTKEINRQKQEHSGHEGGTPRNDGAVALARSOGRVTTDDGYIFNNSDIIEDT	180
DB	122 AHADNVTKEINRQKQEHSGHEGGTSANDGAVAFARSGRYTTDDGYIFNNSDIIEDT	181
QY	181 GDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRYRRONSNTSRTNWVPS	240
DB	182 GDAYIVPHGDHYHYIPKNELASASELAFLSGRENLSNLTYYRRONSNTSRTNWVPS	241
QY	241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQIYKPLSQBVHVESGLVDFDPAQITSR	300
DB	242 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQIYKPLSQBVHVESGLVDFDPAQITSR	301
QY	301 ARGVAVPHGDHYHTPIYQSOMSELEERIARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG	360
DB	302 ARGVAVPHGNHYHTPIYEQMSELEKRIARIIPLYRSNHWVPDSRPEEPSQPTPEPSPS	361
QY	361 PQAPNLK-IDSN---SSLVSQLVKRVKGGYVFEKGLSRVVFPAKOLPSEIVKMLESKLS	416
DB	362 PQAPNPQAPNSNPIDEKLVKEAVKVGKGGYVFEENGVSRYIPAKNLSAETAAGIDSKLA	421
QY	417 KOESVSHLTAKKNVAPRDOEFTDKAYNLTAEKALFYXNGKNSDFOALDKLLERLND	476
DB	422 KOESLSHLGAKKTDLPSSDFEYFNKAYDILLARIHQDLDNKGQVDFEALDNLERLKD	481
QY	477 ESTNKEKLVDDLALFLAPIHTPERLGKPNSEOLEYTEDDEVRIAQADKYTTSDGYIFDHD	536
DB	482 VSSDKVKLVDDLALFLAPIRPERLGKPNSEOLEYTTDDDEIQAVALAGKYTTEDGYIFDPRD	541
QY	537 IISDSDGAVVPHMGHSHWIKGDSLSDEKVAQAAYTKKEGILPSPDADVKANPTGQSA	596
DB	542 IYSDSDGAVVPHMTHSHWIKKDSLSEAEAAAQAYAKEKGLTPSTDHQDSGNTAEKGA	601
QY	597 AAIYNRVKGEKRIPLVRPVPYVHTVFNKGNLLIIPHKDHYHNIKFAFDHHTYKPNGY	656
DB	602 EAIYNRVKAARKVPLDRMPYNLQYTVFNKGNLLIIPHYDHYHNIKFEWDEGLYEAPKGY	661
QY	657 TLEDLPATIKYVVEHPDHPHSDNGWGNASHEVLGKDSHPNKNFKAD-----	706
DB	662 TLEDLIAVTKYVVEHPNERPHSDNGFGNASDHOVQNKNGQATNQTEKPEEKPTQKEPE	721
QY	707 -----EEPVETPAEPVPQVETEKVEAQLKEAEVLLAKVYDSSL	746
DB	722 EETPREEKQSEKPSPKTEPEEESPESEEPQVETEKVEEKLREABDILGKIQDPIT	781
QY	747 KANATETLAGLRNLTQIMDNNSIMAEAKLLALLKGS	795
DB	792 KSNAKETVNGLKNTLFGTDNNTIMAEAEKLLALLKES	820

RESULT 7	
AAB12727	
ID	AAB12727 standard; protein; 821 AA.
XX	
AC	AAB12727;
XX	
DT	21-NOV-2000 (first entry)
XX	

Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA001218.

23-DEC-1998; 98US-0113800P.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N; WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteraemia and/or pneumonia.

Claim 18; Fig 25; 106pp; English.

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-11M protein antigen

Sequence 821 AA;

Query Match 77.3%; Score 3218; DB 3; Length 821;

Best Local Similarity 75.2%; Pred. No. 8.2e-229;

Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

1 SYELGVLQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60

2 AYELGLHQAQTVKNNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 61

61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDXEDIVNEVKGYIVKDGKYYVYLKDA 120

62 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDXEDIVNEVKGYIVKDGKYYVYLKDA 121

121 AHADNVRTKEEINRQKQHSQHQEGGTTPRNDGAVALARSGRYTDDGYIFNASDIIEDT 180

122 AHADNVRTKEEINRQKQHSQHQEGGTTPRNDGAVALARSGRYTDDGYIFNASDIIEDT 181

181 GHAYIVPHGDHYHYIPKNLSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240

182 GHAYIVPHGDHYHYIPKNLSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 241

241 VSNPGTTNTNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 300

242 VSNPGTTNTNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 301

301 ARGVAVPHGDHYHYIPYQMSSELEERARIIPURYNSHNVDPDSRQPSQPTPEPSPG 360

302 ARGVAVPHGNHVFIPYQMSSELEERARIIPURYNSHNVDPDSRQPSQPTPEPSPG 361

361 POPAPNLK-IDSN---SSIVSQLVKRGYVFEKGISRYVFAKDLPSSETVKNLESKLS 416

362 POPAPNPQAPNPIDKLVKAEAVKRGYVFEENGVSRYIPAKNLSAETAGDISKLA 421

417 KOESVSHITLAKENVAPRQGFYQKAYNLLTAHKAFLXNKGNSDFQALDKLLERLND 476

Db 422 KOESLSHLKAGAKKTDLPSSDRFYNNKAYDOLLARIHQDLIDNKGQVDFEALDNLRLKLD 481

QY 477 ESTNKEKLVDDLLAFAPITHPERLCKPNSQIETDEVRIAQLADKYTTSDGYIDEHD 536

Db 482 VSSDKVKLVDDLLAFAPITHPERLCKPNSQIETDEVRIAQLADKYTTSDGYIDEHD 541

QY 537 IISDEGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKKGIPLPSPOADVKANPTGDSA 596

Db 542 ITSDEGDAYVTPHMTSHWIKKDSLSDEKVAQAAYTKKGIPLPSPOADVKANPTGDSA 601

QY 597 AAIYNRVKGEKIPLVRLPYVVEHTVEVKNGLIIPHKDHYHNIKPAFDDHHTYKAPNGY 656

Db 602 EAIYNRVKAAKVPIDRMPYQLQYTVKNGSLIIPHYDHYHNIKFEWDFEGLYKAPNGY 661

QY 657 TLEDLFATIKYVVEHDPDRPHSDNGWNASHLVKDKHSDPDNKNFKAD----- 706

Db 662 TLEDLFATIKYVVEHDPDRPHSDNGWNASHLVKDKHSDPDNKNFKAD----- 706

QY 707 -----BEPVEETPAPEVPOVETEKVEAQLKEAEVLLAKVTDSSL 746

Db 722 BETPREKQSEKPEPKPTPEPESESEPESEPESEPESEPESEPESEPESEPESEPESE 781

QY 747 KANATETLAGLRNLTQLQIMDNNSIMAEAEKLLALLKGS 785

Db 782 KSNKHELTGLKNNLLFGTQDNNTIMAEAEKLLALLKGS 820

RESULT 8

AAU84026

ID AAU84026 standard; peptide; 821 AA.

XX AAU84026;

AC AAU84026;

DT 08-MAY-2002 (first entry)

XX Truncated variant of S. pneumoniae BVH-11, BVH-11M.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; mutein.

OS Streptococcus pneumoniae.

XX Synthetic.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA000908.

XX 20-JUN-2000; 2000US-0212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia.

XX Example 1; Page: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (iii) encoding (i) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (iii) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention

Sequence 821 AA;

Query Match 77.3%; Score 3218; DB 5; Length 821;
Best Local Similarity 75.2%; Pred. No. 8.2e-229;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
2 AVELGLHQAQTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 61
61 SHGDHYHYNGKVPDYAIISELLMKDPNYKLDKEDIVNEVKGVYKVDGKYVYVLKDA 120
62 SHGDHYHYNGKVPDYAIISELLMKDPNYKLDKEDIVNEIKGGVYKVGKYYVYVLKDA 121
121 AHADNVRTKEENNRQKQHSQHREGTTPRNDGAVALARSGQRYTTDDGYIFNASDIIDT 180
122 AHADNVRTKEENNRQKQHSQHREGTSANDGAVAFARSQGRYTTDDGYIFNASDIIDT 181
181 GDAYIVPHGDHYHYIYIPKNLSASELAAEFISGRGNLSNRTYRQNSDNTSRINWVPS 240
182 GDAYIVPHGDHYHYIYIPKNLSASELAAEFISGRGNLSNRTYRQNSDNTSRINWVPS 241
241 VSNPGTNTNTSNNSTNSQASOSNDISLLKQYKLPQSQRHVESDGLVFPQAQITSR 300
242 VSNPGTNTNTSNNSTNSQASOSNDISLLKQYKLPQSQRHVESDGLVFPQAQITSR 301
301 ARGVAVPHGDHYHYIYIPYQMSSELEERARIIPLYRSNHWPDSPRPEQSPQTPPEPSPG 360
302 ARGVAVPHGNHYHYIYIEQMSSELEERARIIPLYRSNHWPDSPRPEQSPQTPPEPSPS 361
361 PQAPNLK-IDSN---SSLVQLVRKVGEGYFEEKGISRYVFAKLPDSETVNLSKLS 416
362 PQAPNPQAPSNPIDEKLVKEAVRKVGDDGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
417 QOESVSHLTAKENVAPRDOEFVDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND 476
422 QOESLGHKGAKTDLDPSSDREFYKAYDILLARIHQDLNKGQVDFEALDNLLELKD 481
477 ESTNKEKLVDDLLAFLAPIHPEPLGKPNQSIETEDVRIAGIADKYTSDDGYIFDEHD 536
482 VSDKVKLVDDLLAFLAPIHPEPLGKPNQITVTDDEIQVAKLAGKYTTEDGYIFDPRD 541
537 IISDEGDAYVTPHNGHSHWTKGSLDKKVAQAQYTKKGLPPSPDADVKANPTGDSA 596
542 ITSDEGDAYVTPHNGHSHWTKGSLDKKVAQAQYTKKGLTPSPDADVKANPTGDSA 601
597 AAIYNRVKGKRIPLVLPYVVEHTVEVGNGLIIPKHQVHNKIFAWFDHHTYKAPNGY 656
602 EAIYNRVKAARKVPLDRMPYLNQITVEVNGSLIIPHYDHYHNKIFEFWDEGLYEA PKGY 661
657 TLEDLFATIKYVVEHPDPRHNSDGNWGNASHVIGKDHSDPDPNNKPAD----- 706
662 TLEDLLATKYVVEHEPERHNSDGNWGNASHVIGKDHSDPDPNNKPAD----- 721
707 -----REPVEETAPBEVQVETEKVEAQLKAEVLLAKVTDSSL 746
722 BETPREEKQSEKPEKPTPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEE 781
747 KANATETIAGLRNNLTQIMDNNSIWAIEKLLALLKGS 785

Db 782 KSNAKETLTGKNNLFGTQDNNTIMAEKLLALLKES 820
RESULT 9
ABM18807
ID ABM18807 standard; protein; 821 AA.
XX AC ABM18807;
XX 13-OCT-2003 (first entry)
XX S. pneumoniae variant protein BVH-11M.
XX antibacterial; antiinflammatory; auditory; vaccine; meningitis;
XX streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;
XX mutant; mutein; New 43; BVH-11; BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO2003054007-A2.
XX 03-JUL-2003.
XX 20-DEC-2002; 2002WO-CA002006.
XX 20-DEC-2001; 2001US-0341252P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
XX Labbe S;
XX WPI; 2003-569224/53.
XX New isolated polypeptides of Streptococcus pneumoniae, useful for
diagnosing, preventing or treating streptococcal infection, meningitis,
otitis media, bacteraemia or pneumonia infection.
XX Example 1; SEQ ID NO 19; 79pp; English.
XX The invention relates to a novel isolated polypeptide of Streptococcus
pneumoniae. A polypeptide of the invention has antibacterial,
antiinflammatory, and auditory activity, and is used as a vaccine. The
polypeptide or pharmaceutical composition is useful for the prophylactic
or therapeutic treatment of streptococcal infection, meningitis, otitis
media, bacteraemia or pneumonia infection. The kit is useful for
detecting or diagnosing streptococcal infection. The pharmaceutical
composition is useful as a vaccine. The polynucleotides are useful in
designing DNA probes for detecting circulating Streptococcus in a
biological sample. The present sequence is used in the exemplification of
the invention. Note: The sequence data for this patent is not fully
represented in the printed specification, but is based on sequence
information supplied by the European Patent Office

Sequence 821 AA;

Query Match 77.3%; Score 3218; DB 7; Length 821;
Best Local Similarity 75.2%; Pred. No. 8.2e-229;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

Qy 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
2 AVELGLHQAQTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 61
61 SHGDHYHYNGKVPDYAIISELLMKDPNYKLDKEDIVNEVKGVYKVDGKYVYVLKDA 120
62 SHGDHYHYNGKVPDYAIISELLMKDPNYKLDKEDIVNEIKGGVYKVGKYYVYVLKDA 121
121 AHADNVRTKEENNRQKQHSQHREGTTPRNDGAVALARSGQRYTTDDGYIFNASDIIDT 180
122 AHADNVRTKEENNRQKQHSQHREGTSANDGAVAFARSQGRYTTDDGYIFNASDIIDT 181

```

QY 181 GDAYIVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 240
DB 182 GDAYIVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 241
QY 241 VSNPGTNTNTSNNSTNSQASQNDIDSLKQLYKPLSQRHVESDGLVDPQAQTSRT 300
DB 242 VSNPGTNTNTSNNSTNSQASQNDIDSLKQLYKPLSQRHVESDGLVDPQAQTSRT 301
QY 301 ARGVAVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 360
DB 302 ARGVAVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 361
QY 361 PQAPNLK-IDSN---SSLVSQLVRKGVGVFEKGIIRYVFAKOLPSEVTKNLESKLS 416
DB 362 PQAPNPQPAPSNPIDKLVKAEVRKVDGVFENGVSRIYPAKNLSAETAAGIDSKLA 421
QY 417 QKESVSHLTAKKNVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 476
DB 422 QKESVSHLTAKKNVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 481
QY 477 ESTNKEKLVDDLLAFAPITHPERLGNKPSQIETDEVRILAQADKYTSDGYIFDEHD 536
DB 482 VSSDKVKLVDDLLAFAPITHPERLGNKPSQIETDEVRILAQADKYTSDGYIFDEHD 541
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 596
DB 542 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 601
QY 597 AAIYNRVKGKRIPLVRLPYMVEHTVFKNGNLIIPHKDHYHNIKFAWDDHTYKAPNGY 656
DB 602 EAIYNRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNIKFEWDEGLYEAPKGY 661
QY 657 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKAD----- 706
DB 662 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKAD----- 721
QY 707 -----EPPVEETPABPEVPQVETEKVEAQLEAVLLAKVTDSSL 746
DB 722 EETPREKPOSEKPEPKTEEPSEESPESEBPQVETEKVEAKLEADLLGKIQDPFI 781
QY 747 KANATETLAGLNRLTIQIMDNNSIMAEKLLALLKGS 785
DB 782 KSNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKES 820

RESULT 10
AAB12716
ID AAB12716 standard; protein; 840 AA.
XX
AC AAB12716;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN W0200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
PR 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

```

```

XX WPI; 2000-452397/39.
DR N-PSDB; AAA65731.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia.
PS Claim 18; Fig 4; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents the S. pneumoniae BVH-11
CC protein antigen
XX
SQ Sequence 840 AA;
Query Match 77.3%; Score 3218; DB 3; Length 840;
Best Local Similarity 75.2%; Pred. No. 8.5e-229;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
QY 1 SYELGLQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
DB 21 AYLGLHQAQTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKEDIIVNEVKGQYIVTKDQKYYVYLKDA 120
DB 81 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKEDIIVNEVKGQYIVTKDQKYYVYLKDA 140
QY 121 AHADNVRTKEEINRQKQESQHREGGTPRNDGVALARSQGYTTDDGYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQESQHREGGTSNDGAVAFARSQGYTTDDGYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 240
DB 201 GDAYIVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 260
QY 241 VSNPGTNTNTSNNSTNSQASQNDIDSLKQLYKPLSQRHVESDGLVDPQAQTSRT 300
DB 261 VSNPGTNTNTSNNSTNSQASQNDIDSLKQLYKPLSQRHVESDGLVDPQAQTSRT 320
QY 301 ARGVAVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 360
DB 321 ARGVAVPHGDHYHYPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRNTNWVPS 380
QY 361 PQAPNLK-IDSN---SSLVSQLVRKGVGVFEKGIIRYVFAKOLPSEVTKNLESKLS 416
DB 381 PQAPNPQPAPSNPIDKLVKAEVRKVDGVFENGVSRIYPAKNLSAETAAGIDSKLA 440
QY 417 QKESVSHLTAKKNVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 476
DB 441 QKESVSHLTAKKNVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 500
QY 477 ESTNKEKLVDDLLAFAPITHPERLGNKPSQIETDEVRILAQADKYTSDGYIFDEHD 536
DB 501 VSSDKVKLVDDLLAFAPITHPERLGNKPSQIETDEVRILAQADKYTSDGYIFDEHD 560
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 596
DB 561 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 620
QY 597 AAIYNRVKGKRIPLVRLPYMVEHTVFKNGNLIIPHKDHYHNIKFAWDDHTYKAPNGY 656
DB 621 EAIYNRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNIKFEWDEGLYEAPKGY 680
QY 657 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKAD----- 706
DB 681 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKAD----- 740

```

QY	707	-----EEPVETPAPEVPQVETEKVEAQIKBAEVLAKVTDSSL	746
DB	741	EEPEEEPEEPOSEKPEEPPEESPESEEPVETEKVEKUREAEDLGIQDPII	800

QY 747 KANATETLAGLRNNLTQIMDNNSIMAAEAKLLALLKGS 785
| : | | | | : | | | | : | | | | |
Dh 801 KSNAKETTGLKNLLFGTDNNTIMAAEAKLLALKES 839

RESULT 11

AAU75933
ID AAU75933 standard; protein; 840 AA.

AA
AC AAU75933:XX
DT 08-MAY-2011XX
DE
Streptococcus pneumoniae BXX
XX
XX
RVH-3: RVH-11: vaccine: meningitis: otitis

KW streptococcal bacterial infection.
yy

OS Streptococcus pneumoniae.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA000908.

PR 20-JUN-2000; 2000US-0212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

DR N-PSDB; ABK13103.
XX

PT new streptococcus pneumoniae ssn-3 and ssn-2 vaccines
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteremia.

XX
PS
Example 1: Fig 7: 113pp: English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g. caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This is the amino acid sequence of Streptococcus pneumoniae protein BVH-11, used to create the antigenic peptides described in the method of the invention.

XX
Sequence 840 AA:

Query Match 77.3%: Score 3218; DB 5; Length 840;

Query Match	77.3%	Score 3218	22.5%	Length 310
Best Local Similarity	75.2%	Pred. No. 8.5e-229		
Matches 616; Conservative	65	Mismatches 104	Indels	34; Gaps 3;

QY	1	SYELGLOYASTVKKENNRSVYIDGKQATQKTENLTPEDEVSKREGINABQIVIKITDQGYVT	60
Db	21	AYELGULOQATVTKENNRSVYIDGKQATQKTENLTPEDEVSKREGINABQIVIKITDQGYVT	80
QY	61	SHGDHYHYNGKVPYDAIIISELLMKDPNPKLKDEDIVNEVKGYIVTKDGKYYVYLKDA	120
Db	81	SHGDHYHYNGKVPYDAIIISELLMKDPNQLKDSDIVNEIKGGYIVTKVNGKYYVYLKDA	140
QY	121	AHADNVTKBEINRQKQEHSHQHEGGTPRNDGAVALARSGRYTTDGGYIFNASDIIDET	180
Db	141	AHADNVTKBEINRQKQEHSHQHEGGTSANDGAVAFARSGRYTTDGGYIFNASDIIDET	200
QY	181	GDVIVPHGDHYHYIPKNELASLAAAEAFLSGRGLNSRSTRYRRQNSDNTSRTNWVPS	240
Db	201	GDVIVPHGDHYHYIPKNELASLAAAEAFLSGRENLSNLTYYRRQNSDNTPRTNWVPS	260
QY	241	VSNPGTTNTNTSNNSNTNSQASQNDIDSLIKQIKYKPLSORHVESDGLVPDPAQITSR	300
Db	261	VSNPGTTNTNTSNNSNTNSQASQNDIDSLIKQIKYKPLSORHVESDGLVPDPAQITSR	320
QY	301	ARGVAVPHGDHYHYIPYQMSLEERITARIIPLYRSNHNWVPDGRPEQSPQPTPEPSPG	360
Db	321	ARGVAVPHGNHYHIPYQMSLEKRTARIIPLYRSNHNWVPDGRPEQSPQPTPEPSPS	380
QY	361	PQAPNLK-IDSN---SSLVSQLVKRVGEGVYVPEEKGISRVVFAKDIIPSETVKMLSEKLS	416
Db	381	PQAPNPQAPSNPIDEKLVKEAVKRVGDGVVPEENGVSRYIPAKNISAEATAAGIDSKLA	440
QY	417	KQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLND	476
Db	441	KQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLDNKGQVDEALDNLLEKLD	500
QY	477	ESTNKEKLVDDLLAFAPITHEPRIGKPNISOIETEDEVRIAQLADKYTTSDGYIFDSDH	536
Db	501	VSSDKVKLVDDLLAFAPIRHEPRIGKPNQITVTDDBIQVAKLAGKYTTEDGYIFDPRD	560
QY	537	IISDEGDVYTPHMGHSHWIKGDSLSDEKEVAAQAYTKEKGLPPSPADVKANPTGDSA	596
Db	561	ITSDEGDVYTPHMTSHWIKKDSLSSEAEAAAQAYAKEKGLTPSTDHQDSGNTAEKGA	620
QY	597	AAIYNVRVGEKRIPLVRIPYVVEHTVEVKNGNLIIPHDXDHVNIKFAMFDDHHTYKAPNGY	656
Db	621	EAIYNVRAAKVPLDRIPYVLYQYIVVEVNGSLIIPHVDYHNIKEFWFDEGLYEAPGY	680
QY	657	TLEDLFAITKYVEHPDRPHSNDGWSAEHVILKGDHSDPNKNFKAD-----	706
Db	681	TLEDLLATVKYVVEHPNRPNSHDNGFNASDHYVQRNKGQADTNQTEKPEEKPQTEKPE	740
QY	707	-----REPVEETAEPPEVPOVETEKVEAQKAEAVILLAKVTDSSL	746
Db	741	BETPREKPSQEKPSPKPTPEEESPEESEPQVETEKVEEKUREADELLGKIQDPPI	800
QY	747	KANATETLAGLRNNLTQLMDNNSTMAEAKLLALLKGS	785
Db	801	KSNAKETLTGLKNNLLFGTQDNNITMAEAKLLALLKES	839
RESULT 12			
ABM18797			
ID	ABM18797	standard; protein; 840 AA.	
XX	AC	ABM18797;	
XX	XX		
DT	13-OCT-2003	(first entry)	
DE	DE	S. pneumoniae BVH-11 polypeptide SEQ ID NO: 8.	
XX	XX		
KW	KW	antibacterial; antiinflammatory; auditory; vaccine; meningitis;	
KW	KW	streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3.	
OS	OS	Streptococcus pneumoniae.	
XX	XX		
PN	PN	WO2003054007-A2.	

[illegible]

444

Db 422 KQESLSHKLGAKTDLPSDDREFYKAYDILLARHQDLNDKGRQVDFEALDNLRLERKD 481
 QY 477 ESTNKEKLVDDLLAFAPITHTPERLGKPNQSEIYETDEVRIAQLADKYTTSDGYIFDEHD 536
 Db 482 VSDKVKLVDDLLAFAPITHTPERLGKPNQSEIYETDEVRIAQLADKYTTSDGYIFDEHD 541
 QY 537 IISDEGDAYVTPHMGSHWIGKSDSKDEKVAQAQYTKKGLPSPDADVKANPTGDSA 596
 Db 542 ITSDEGDAYVTPHMGSHWIKKSDLSAEARAAQAYAKEGLTPSTDHDSGNTAEAKGA 601
 QY 597 AAIYNRVKEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNKFAWDDHTYKAPNGY 656
 Db 602 EAIYNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNKFEWFEGLYEAPKGY 661
 QY 657 TLEDLFATIKYVVEHDPDRPHSDNGWNA 685
 Db 662 TLEDLLATVKYVVEHPNRPSPDNGFGNA 690

RESULT 15
 ABM18826
 ID ABM18826 standard; protein; 690 AA.
 AC ABM18826;
 XX
 DT 13-OCT-2003 (first entry)
 DE S. pneumoniae variant protein NEW16.
 KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
 KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;
 KW mutant; mutein; New 43; BVH-11; BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 OS
 PN WO2003054007-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-CA002006.
 XX
 PR 20-DEC-2001; 2001US-0341252P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
 PI Labbe S;
 XX
 DR WPI; 2003-569224/53.
 XX
 PT New isolated polypeptides of Streptococcus pneumoniae, useful for
 PT diagnosing, preventing or treating streptococcal infection, meningitis,
 PT otitis media, bacteraemia or pneumonia infection.
 XX
 PS Example 1; SEQ ID NO 38; 79pp; English.
 XX

The invention relates to a novel isolated polypeptide of Streptococcus pneumoniae. A polypeptide of the invention has antibacterial, antiinflammatory, and auditory activity, and is used as a vaccine. The polypeptide or pharmaceutical composition is useful for the prophylactic or therapeutic treatment of streptococcal infection, meningitis, otitis media, bacteraemia or pneumonia infection. The kit is useful for detecting or diagnosing streptococcal infection. The pharmaceutical composition is useful as a vaccine. The polynucleotides are useful in designing DNA probes for detecting circulating Streptococcus in a biological sample. The present sequence is used in the exemplification of the invention. Note: The sequence data for this patent is not fully represented in the printed specification, but is based on sequence information supplied by the European Patent Office

Sequence 690 AA;
 SQ

Query Match 71.1%; Score 2961; DB 7; Length 690;
 Best Local Similarity 80.7%; Pred. No. 6.4e-210;
 Matches 556; Conservative 54; Mismatches 75; Indels 4; Gaps 2;
 QY 1 SYELGLYQARTVKENRVSIDGKQATOKTENLTDEVSKREGINAEQIVIKITDQGYVT 60
 Db 2 AVELGLHQATVKENRVSIDGKQATOKTENLTDEVSKREGINAEQIVIKITDQGYVT 61
 QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIDVNEVKGVIKVDGKYVYLKDA 120
 Db 62 SHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIDVNEIKGGYVKNKGKYVYLKDA 121
 QY 121 AHADNVRTKEINRQKQHSQHREGGTPRNGGAVALARSQGYTTDDGYIFNASDIIEET 180
 Db 122 AHADNVRTKEINRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASDIIEET 181
 QY 181 GDAYIVPHGDHYHYIPKKNELASASLAAAEAFLSGRGNLSNSTRYRQNSDNTSRNWPVS 240
 Db 182 GDAYIVPHGDHYHYIPKKNELASASLAAAEAFLSGRGNLSNSTRYRQNSDNTSRNWPVS 241
 QY 241 VSNPTNTNTSNNSNTNSQASQNSNDISLLKQLYKLPISQRHVESDGLVDFPAQITSR 300
 Db 242 VSNPTNTNTSNNSNTNSQASQNSNDISLLKQLYKLPISQRHVESDGLVDFPAQITSR 301
 QY 301 ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
 Db 302 ARGVAVPHGNHYHFIPEQMSSELEKRIARIIPLYRSNHWVPDSRPEPSPQPTPEPSPS 361
 QY 361 POPAPNLK-IDSN--SSLSYQLYKVGEGYVFEKGLSRVYFAKDLPSSETVKNLESKLS 416
 Db 362 POPAPNQPPAPSNPIDKLVKAEARKVKGVDGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
 QY 417 KQESVSHTLTAKENVAAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFOALDKLLERLND 476
 Db 422 KQESLSHKLGAKTDLPSDDREFYKAYDILLARHQDLNDKGRQVDFEALDNLRLERKD 481
 QY 477 ESTNKEKLVDDLLAFAPITHTPERLGKPNQSEIYETDEVRIAQLADKYTTSDGYIFDEHD 536
 Db 482 VSDKVKLVDDLLAFAPITHTPERLGKPNQSEIYETDEVRIAQLADKYTTSDGYIFDEHD 541
 QY 537 IISDEGDAYVTPHMGSHWIGKSDSKDEKVAQAQYTKKGLPSPDADVKANPTGDSA 596
 Db 542 ITSDEGDAYVTPHMGSHWIKKSDLSAEARAAQAYAKEGLTPSTDHDSGNTAEAKGA 601
 QY 597 AAIYNRVKEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNKFAWDDHTYKAPNGY 656
 Db 602 EAIYNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNKFEWFEGLYEAPKGY 661
 QY 657 TLEDLFATIKYVVEHDPDRPHSDNGWNA 685
 Db 662 TLEDLLATVKYVVEHPNRPSPDNGFGNA 690

RESULT 16

AAAY91939
 ID AAAY91939 standard; protein; 826 AA.
 XX

AC AAAY91939;
 XX

DT 19-JUL-2000 (first entry)
 XX

DE S. pneumoniae 92 kDa human C3-degrading protein.
 XX

KW Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
 KW inhibitor; inflammation; organ rejection; xenotransplantation.
 XX

OS Streptococcus pneumoniae.
 XX

PN WO200017370-A1.
 XX

PD 30-MAR-2000.
 XX

PF 24-SEP-1999; 99WO-US022362.
 XX

XX	24-SEP-1998;	98US-0101736P.	Qy	536	DIISDEGDVYTPHMGHSHWIGKDSLSDEKVAQAQYTKKGLPPSPDADVKANPTGDS	595
PR	31-MAR-1999;	99US-00283094.	Db	547	DITSDEGDVYTPHMGHSHWIKDLSGEAERAAQAQYAKKGLTPPSTHDODSGNTRAKG	606
XX	(MINU) UNIV MINNESOTA.		Qy	596	AAAIYRVKGEKRIPLARLPMVVEHTVEVKNGLIIPKDHVHNKIPAFWDDHTYKAPNG	655
PA	(AMCY) AMERICAN CYANAMID CO.		Db	607	AAAIYRVKAAKVPDLRMPYNLQYTVVKNGLIIPHYDHYHNKIPFWDGLYEAPKG	666
XX	Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;		Qy	656	YTLDELATIKYVVEHPDERPHSDGNGMGNASEHVLGKDKHSDSPNKNFKADEE-----P	709
XX	WPI; 2000-283594/24.		Db	667	YTLDELATIKYVVEHPDERPHSDGNGMGNASEHVLGKDKHSDSPNKNFKADEE-----P	726
XX	N-PSDB; AAA08557.		Qy	710	VEET-----PABEVEVQVETEKVEAQIKAEVLLAKVTDSSL	746
XX	Isolated polypeptide is used to stimulate immune system and immunize or		Db	727	EETPREKPKQSEKPEPKTEEPSEPESEEPQVETEKVEKUREAEDLLGKIQDPII	786
XX	colonization.		Qy	747	KANATETLAGLNLTQIMDNNSIMAEKLLALLKGS	785
XX	Claim 8; Page 55-57; 63pp; English.		Db	787	KSNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKES	825
XX	The present sequence, isolated from Streptococcus pneumoniae, is a human		XX	RESULT 17		
CC	C3-degrading protein of about 92 kDa. This sequence may encompass a		XX	ASU01597		
CC	smaller, approximately 20 kDa protein (see AA919198), also having human		ID	ABU01597 standard; protein; 819 AA.		
CC	C3-degrading activity. The DNA sequences (AAA08556-57) can be used for		AC	ABU01597;		
CC	producing an immune response to Streptococcus pneumoniae in a mammal.		XX	23-OCT-2003 (revised)		
CC	Antibodies against the proteins can be used to inhibit S. pneumoniae-		DT	11-FEB-2003 (first entry)		
CC	mediated C3 degradation. C3-mediated inflammation and rejection in		XX	S. pneumoniae type 4 strain protein from coding region #1173.		
CC	xenotransplantation can be inhibited by expressing the nucleic acid		XX	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;		
CC	sequences on the surface of an organ of an animal. In particular, the		KW	ant inflammatory; antibacterial; immunostimulant; auditory; respiratory;		
CC	polypeptides are useful for stimulating the immune system and are		KW	gene therapy; vaccine.		
CC	effective to immunize or treat a mammalian subject against Streptococcus		OS	Streptococcus pneumoniae; type 4 strain.		
CC	pneumoniae infection or colonization		XX	WO200277021-A2.		
XX	Sequence 826 AA;		XX	03-OCT-2002.		
SQ	Query Match 67.5%; Score 2812; DB 3; Length 826;		XX	27-MAR-2002; 2002WO-IB002163.		
	Best Local Similarity 67.4%; Pred. No. 9e-199;		XX	27-MAR-2001; 2001GB-00007658.		
	Matches 552; Conservative 81; Mismatches 138; Indels 48; Gaps 8;		XX	(CHIR-) CHIRON SPA.		
			XX	(GENO-) INST GENOMIC RES.		
Qy	1 SYELGLYQARTV-KENNRVSYIDGKAOTKENTLTPDEVSKREGINAEQIVIKITDQGVV	59	PI	Masignani V, Tettelin H, Fraser C;		
Db	21 SYELGRHQAGQVKESNRVSYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQGVV	80	XX	WPI; 2003-040579/03.		
Qy	60 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIYNEIKGGYVIVKVGKYYVYVKD	119	DR	N-PSDB; ABX06885.		
Db	81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIYNEIKGGYVIVKVGKYYVYVKD	140	XX	New proteins and nucleic acid molecules from Streptococcus pneumoniae,		
Qy	120 AAHADNVRTKEINRQKQHSQHRGEGTPNDGAVALARSGHYTTDDGYIFNASDIIE	179	PT	useful as medicaments for treating or preventing a disease or infection		
Db	141 AAHADNVRTKEINRQKQHSQHRGEGTPNDGAVALARSGHYTTDDGYIFNASDIIE	198	PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or		
Qy	180 TGDAYIVPHGDHYHYTPKNELSAELAAEFLSGRNLNSRTYRRQNSDNTSRTNWVP	239	XX	ear infection.		
Db	199 TGDAYIVPHGDHYHYTPKNELSAELAAEFLSGRNLNSRTYRRQNSDNTSRTNWVP	247	PS	Claim 1; SEQ ID NO 2346; 56pp; English.		
Qy	240 SVNPGTNTNTSNTSNTSQAQSDNDIDLLKQLYKPLSRHVESDGLVFPQAQITSR	299	CC	The invention relates to a protein comprising or having at least 50%		
Db	248 NPAQPLSENHNTVTPYHQ-NQGENISLLRELYAKPLSERHVESDGLVFPQAQITSR	306	CC	identity to any of the 2469 amino acid sequences, identified in the		
Qy	300 TARGVAVPHGDHYHYTPYQMSLEERARIIPLRVRSNHWVPDSRPEQSPQPTPEPSP	359	CC	specification (available on a computer readable format), or its fragment,		
Db	307 TARGVAVPHGDHYHYTPYQMSLEERARIIPLRVRSNHWVPDSRPEQSPQPTPEPSP	366	CC	expressed from 2469 of 2489 identified DNA coding regions from the		
Qy	360 GPOAPNLK-IDSN---SSIVSQVLRVGVGVYFEEKIGISYVFAKPLSETVKNLESKL	415	CC	Streptococcus pneumoniae type 4 strain genomic sequence appearing as		
Db	367 SPOAPNPQAPSNPIDEKLVKAVRKGVGGVYFENGVSRYIPAKDLSAETAAGIDSKL	426	CC	AS956454. Also included are an antibody which binds one of the proteins,		
Qy	416 SKQESVSHLTAKENVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLRLN	475	CC	treatment of a patient by administering the protein, DNA or antibody (in a		
Db	427 AKQESLSHKLGAKKTDLPSSDRFPYKAYDILLARIHQDLDDNKGQVDFEALDNLRLK	486	CC	nucleic acid cited above or fragments between nucleotides 8-100 of a		
Qy	476 DESTNKEKLVDDLAFLAPIHTHPRKLPNSQIETDEVRVIAQLADKYTSDGYIDEH	535	CC	sequence not defined in the specification, for amplifying a target		
Db	487 DVPSDKVKLVDDLAFLAPIHTHPRKLPNSQIETDEVRVIAQLADKYTSDGYIDEH	546	CC	sequence contained within a Streptococcus nucleic acid sequence, where		
			CC	the first primer is substantially complementary to the target sequence		

CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)

XX Sequence 819 AA;

Query Match 67.2%; Score 2797.5; DB 6; Length 819;
 Best Local Similarity 67.0%; Pred. No. 1.1e-197;
 Matches 546; Conservative 84; Mismatches 136; Indels 47; Gaps 7;

QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYV 59
 Db 21 SYELGRYQAGQDKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIIVNEKGYVIVKDPKYYVYVKD 119
 Db 81 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIIVNEKGYVIVKDPKYYVYVKD 140
 QY 120 AAHADNVRTKEEINRQKQESHREGGTPRNDGVALARSQRYTTDDGYIFNADIIED 179
 Db 141 AAHADNVRTKEEINRQKQESHREGGTPRNDGVALARSQRYTTDDGYIFNADIIED 197
 QY 180 TGDAYIVPHGDHYHYIPKNELASSELAAAEFLSGRGNLSNRTYRQNSDTSRTNWP 239
 Db 198 TGDAYIVPHGDHYHYIPKNELASSELAAAEFLSGRGNLSNRTYRQNSDTSRTNWP 246
 QY 240 SVSNPGTNTNTSNNSNTNSQASNDIDSLKOLYKPLSORHVESDGLVFPQAITSR 299
 Db 247 NPQAPRLSENHLNLTPTTYHQ-NQGENISSLLRELYAKPLSERHVESDGLVFPQAITSR 305
 QY 300 TARGVAVPHGDHYHYFIPYSQMSLEERARTIPLRYESNHWVDSRPEQSPQPTPEPSP 359
 Db 306 TARGVAVPHGDHYHYFIPYSQMSLEERARTIPLRYESNHWVDSRPEQSPQPTPEPSP 365
 QY 360 GPQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSQKE 419
 Db 366 SPQAPNSPID--EKLKVAEVRKVGDCGVFEENGVSRYIPAKDLSATAAGIDSKLAKQE 423
 QY 420 SVSHHTLTAKENAVPRDQEFYKAYNLLTAHKAFLFNKGRNSDFQALDKLERLNDEST 479
 Db 424 SLHGKLTGKTKTDLPSSDREYFNKAYDILLARIHQDLNDKGRQVDFEALDNLRLKDVSS 483
 QY 480 NKEKLVDDLLAFAPLTHPRLKPNISOIEYTVDEVRILQADKYTTSDGYIFDEHDIIS 539
 Db 484 DKVKLVEDILAFAPLTHPRLKPNISOIEYTVDEVRILQADKYTTSDGYIFDEHDIIS 543
 QY 540 DEGDVAVTPHMGHSHWIGKDSLSKQKVAQAAYTKBKGLPSPDADVKANPTGDSAAAI 599
 Db 544 DEGDVAVTPHMGHSHWIGKDSLSKQKVAQAAYTKBKGLPSPDADVKANPTGDSAAAI 603
 QY 600 YNRVKEKRIPLVRLPYMVEHTVEVNGNLIIPHDKHYHNKIFAWFDDHTYKAPNGYTLE 659
 Db 604 YNRVKAACKVPLDRMPYNLYQYTVVEVNGSLIIPHYDHYHNKIFAWFDDHTYKAPNGYTLE 663
 QY 660 DLFAITKYVVEHPDERPHNDGNGNASEHVLGKDHSEDPNKNFKADEE-----PVEET 713
 Db 664 DLLATVYVVEHPDERPHNDGNGNASEHVLGKDHSEDPNKNFKADEE-----PVEET 723

QY 714 -----PAPPEVPQVETEKVAQLKEAEVLLAKVTDSSLKANA 750
 Db 724 PREEKQSEKPSPKPTPEEPSESESPQVETEKVEKLEAEADLLCKIQDPIKNSA 783
 QY 751 TETLAGRLNLLTQIMDNNSINAEAKKLLALLKGS 785
 Db 784 KETLGLKNLLFGTDQNTNIMAEAKKLLALLKES 818
 RESULT 18
 AAY81662
 ID AAY81662 standard; protein; 827 AA.
 AC AAY81662;
 XX
 DT 24-MAY-2000 (first entry)
 XX
 DE Streptococcus pneumoniae protein sequence ID311.
 XX
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200006737-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB002451.
 XX
 PR 27-JUL-1998; 98GB-00016337.
 PR 19-MAR-1999; 99US-0125164P.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Gilbert CFG, Hansbro PM;
 XX
 DR MPI; 2000-195300/17.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing or
 PT inhibiting expression of the protein.
 XX
 PS Claim 2; Page 99-100; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties. The
 CC protein sequences, and fragments of them, are useful as immunogens and/or
 CC antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention
 XX
 SQ Sequence 827 AA;

Query Match 67.1%; Score 2795; DB 3; Length 827;
 Best Local Similarity 67.0%; Pred. No. 1.1e-197;
 Matches 549; Conservative 83; Mismatches 139; Indels 48; Gaps 8;

QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYV 59
 Db 21 SYELGRHQAQDKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIIVNEKGYVIVKDPKYYVYVKD 119
 Db 81 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIIVNEKGYVIVKDPKYYVYVKD 140

```

PR 21-DEC-1998; 98US-0113048P.
XX (MEDI-) MEDIMMUNE INC.
PA Johnson LS, Koenig S, Adamou JE;
XX WPI; 2000-452129/39.
XX N-PSDB; AAA47605.
XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX such as otitis media, nasopharyngeal and bronchial infections, comprises
XX Streptococcus pneumoniae proteins.
XX Claim 1; Page 65-69; 70pp; English.
XX
XX Although a number of proteins have been suggested as being involved in
XX the pathogenicity of Streptococcus pneumoniae, there still remains a need
XX to identify polypeptides having epitopes in common from various strains
XX of S. pneumoniae in order to utilize such polypeptides in vaccines to
XX protect against a wide variety of S. pneumoniae. New vaccine compositions
XX are described which comprise a Streptococcus pneumoniae polypeptide (or
XX fragments) of 80 - 680 amino acids in length that comprise at least one
XX histidine triad residue (HxxHxx) or a coiled-coil region, or an antibody
XX directed against these features. The vaccine is useful in protecting
XX against infection by Streptococcus pneumoniae. The vaccine composition
XX comprising antibodies to is useful for passive immunization for treating
XX pneumococcal infections which includes otitis media, nasopharyngeal and
XX bronchial infections
XX
XX Sequence 819 AA;
XX
XX Query Match 67.0%; Score 2788.5; DB 3; Length 819;
XX Best Local Similarity 66.7%; Pred. No. 4.9e-197;
XX Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;
XX
XX QY 1 SYELGLYQA-RIVKNNRVSYIDGQAKTKENLTPDEVSKREGINAEOIVIKITDQGV 59
XX Db 21 SYELGRYQAGQDKESNRVAYIDGQAGQAKENLTPDEVSKREGINAEOIVIKITDQGV 80
XX QY 60 TSHGDHYHYNGKVPYDAIISBELMKDPNYKLDIEDIVNEKGGYVIKVGKYYVYKLD 119
XX Db 81 TSHGDHYHYNGKVPYDAIISBELMKDPNYKLDIEDIVNEKGGYVIKVGKYYVYKLD 140
XX QY 120 AAHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIED 179
XX Db 141 AAHADNVRTKEIKRQKQSHNHS--RADNAVAARAAQGRYTTDDGYIFNASDIIED 197
XX QY 180 TGDYIVPHGDHYHYIPKNELSASELAARAEFLSGRGNLSNRTYRQNSDNTSRTNWVP 239
XX Db 198 TGDYIVPHGDHYHYIPKNELSASELAARAEFLSGRGNLSNRTYRQNSDNTSRTNWVP 246
XX QY 240 SVSNPGTTNTNTSNNTSNQASQSDNDISLLKQLKPLSQRHVESDGLVFPDPAQITSR 299
XX Db 247 NPAQPLSENHNLTVPTTYHQ-NQGENISSLLRELYAKPLSERHVESDGLVFPDPAQITSR 305
XX QY 300 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVDSPREPSPQTPPEPSP 359
XX Db 306 TARGVAVPHGNHYHYIPYEQMSELEKRIARIIPLYRSNHWVDSPREPSPQTPPEPSP 365
XX QY 360 GPQAPNLIKDSNLSVQLVRKVGEGYFEEKIGISRYVFAKDLPSSETVKNLESKSKQE 419
XX Db 366 SPQAPFNSNP.D--GKLVKEAVRKVGEGYFEEKIGISRYVFAKDLPSSETVKNLESKSKQE 423
XX QY 420 SVSHLTITAKENVAPRQDEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDST 479
XX Db 424 SLSHKLGTKTKDLPSSDREFFYNKAYDILLARIHQDLDLNDKGRQVDFEALDNLRLKDVSS 483
XX QY 480 NKEKLVDDLLAFAPITHPERLKPNSQIEYDEVRIAQLADKYTTSDGYIFDEHDIIS 539
XX Db 484 DKVKLVDEILAFAPITHPERLKPNSQIEYDEVRIAQLADKYTTSDGYIFDEHDIIS 543
XX QY 540 DEGDAYVTPHMGSHWIGKDSLDEKVAQAAYTKEGILPPSPDADVKANPTGDSAAA 599
XX
XX RESULT 19
XX AAB01469
XX ID AAB01469 standard; protein; 819 AA.
XX AC AAB01469;
XX XX
XX XX 20-OCT-2000 (first entry)
XX DT
XX DE Recombinant variant of Sp36 (Sp36B) of S. pneumoniae.
XX KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
XX KW histidine triad residue; Sp36; antibody; otitis media;
XX KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
XX KW meningitis; lobar pneumonia.
XX OS
XX OS Streptococcus pneumoniae.
XX XX
XX PN W0200037105-A2.
XX XX
XX PD 29-JUN-2000.
XX PF
XX PF 21-DEC-1999; 99WO-US030390.
XX XX

```

Db 544 DEGDAYVTPHMTSHWIKKDSLSAEARAAQAAYAEKGLTPPSTDHODSGNTEAKGAEAI 603
 QY 600 YNRVKGKRIPIPLRVPMVEHTVEKNGNLIIPHKHVHNKIFKAFDDHTYKAPNGYTYLE 659
 Db 604 YNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNKIFEFWDEGLYEAPKGYTYLE 663
 QY 660 DLFATIKYVVEHPDERPHSDNGWGNASEHVLGKDKHSDPNKFNKADDE-----PVEET 713
 Db 664 DLLATVKYVVEHPNERPHSDNGFGNASDHVQKNKGQADTNQTEKPSSEKQTEKPEET 723
 QY 714 -----PAPEVPOVETEKVEAQKAEVLLAKVTDSSLKANA 750
 Db 724 PREEKQSEKPEPKPTPEEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 783
 QY 751 TETLAGLNNTLTQIMDMNNSIMAEKLLALLKGS 785
 Db 784 KETLTGLKNLLFGTDQNNTIMAEKLLALLKES 818

RESULT 20
 AAB12764
 ID AAB12764 standard; protein; 805 AA.
 AC AAB12764;
 XX
 XX 21-NOV-2000 (first entry)
 DT Streptococcus pneumoniae strain JNR7/87 BVH-11 protein antigen.
 DE
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX Streptococcus pneumoniae.
 OS
 OS WO200039299-A2.
 PN 06-JUL-2000.
 PD
 XX 20-DEC-1999; 99WO-CA001218.
 XX
 XX 23-DEC-1998; 98US-0113800P.
 PR (BIOC-) BIOCHEM PHARMA INC.
 PA
 PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI WPI; 2000-452397/39.
 XX
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 PS Disclosure; Fig 12; 106pp; English.
 XX

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, CC may be used for the recombinant production of the proteins they encode. CC The protein antigens may then be used as vaccines for the prevention and CC treatment of Streptococcal infections in mammals (especially humans) CC which result in, e.g. meningitis, otitis media, bacteraemia and/or CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein CC antigen, from the present invention
 XX
 SQ Sequence 805 AA;
 Query Match 66.9%; Score 2786; DB 3; Length 805;
 Best Local Similarity 66.3%; Pred. No. 7.3e-197;
 Matches 544; Conservative 88; Mismatches 136; Indels 52; Gaps 7;
 QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATKENTLTPDEVSKREGINAEQIVIKITDQGV 59
 Db 2 SYELGRHQAGQDKKESNRVAYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQGV 61

QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGYVIVKDGKYYVYLKD 119
 Db 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEIKGGYVIVKNGKYYVYLKD 121
 QY 120 AAHADNVRTKEEINROKQEHSHQREGGTPRNDGVALARSQGRVTTDDGYIFNASDIIED 179
 Db 122 AAHADNVRTKEEIKRQKQERSHNHNS---RADNVAARAQGRYTTDDGYIFNASDIIED 178
 QY 180 TGDYIIVPHGDHYHYIPKNELASASELAASAEFLSGRGNLSNRTVRRNSONTSRNTNVP 239
 Db 179 TGDYIIVPHGDHYHYIPKNELASASELAASAEAYWNG-----KQSPSPSSSSSYNA 227
 QY 240 SVSNPGTNTNTNSNTNSQASQSDNDLSLLKOLYKPLSORHVESDGLVDFDPAQITSR 299
 Db 228 NPAQFRLSENHNLTVTPTYHQ-NQGENISLLRELYAKPLSERHVESDGLVDFDPAQITSR 286
 QY 300 TARGVAVPHGDHYHYFIPYSOMSELEERTIARIPIRYRSHNHWVPSRPEQSPQTPPEPSP 359
 Db 287 TARGVAVPHGNHYHYFIPYEQNSELEKRIARIPIRYRSHNHWVPSRPEPSPQTPPEPSP 346
 QY 360 GPQAPNLKIDSSSLVSQVLRKVGEGYVPEEKGISRYVFAKDLPSSETVKNLEKLSKOE 419
 Db 347 SPQAPNSPID-EKLVKEAVRKVGDSYVFEENGVSRYIIPAKDLSAETAAGIDSKLAKOE 404
 QY 420 SVSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDEST 479
 Db 405 SLSHKLGAKKTDLPSSDRFYNKAYDILLARIHQDLNDKGRQVDFEALDNLRLKDVSS 464
 QY 480 NKEKLVDDLLAFAPITHPERLGPKNISOBYTEDEVRIAGIADKYTTSDGYIFDEHDIIS 539
 Db 465 DKVCLVDIIILAFAPINHPERLGPKNQAITYTDDIQAIVAKLAGKYTTEDGYIFDPRDITS 524
 QY 540 DEGDAYVTPHMGHSHWIKKDSLSDEKKEVAAQAAYTKEGILPPSPDADVKANPTGDSAAAI 599
 Db 525 DEGDAYVTPHMTSHWIKKDSLSAEARAAQAAYAEKGLTPPSTDHODSGNTEAKGAEAI 584
 QY 600 YNRVKGKRIPIPLRVPMVEHTVEKNGNLIIPHKHVHNKIFKAFDDHTYKAPNGYTYLE 659
 Db 585 YNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNKIFEFWDEGLYEAPKGYTYLE 644
 QY 660 DLFATIKYVVEHPDERPHSDNGWGNASEHVLGKDKHSDPNKFNKADDE-----PVEET 713
 Db 645 DLLATVKYVVEHPNERPHSDNGFGNASDHVQKNKGQADTNQTEKPNNEEKQTEKPEET 704
 QY 714 -----PAPEVPOVETEKVEAQKAEVLLAKVTDSS 745
 Db 705 PREEKQSEKPEPKPTPEEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 764
 QY 746 LKANATETLAGLNNTLTQIMDMNNSIMAEKLLALLKGS 785
 Db 765 IKSNAKETLTGLKNLLFGTDQNNTIMAEKLLALLKES 804

RESULT 21

AAB12755

ID AAB12755 standard; protein; 820 AA.

XX

AC AAB12755;

XX

DT 21-NOV-2000 (first entry)

XX

DE Streptococcus pneumoniae strain JNR7/87 BVH-11-2 protein antigen.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

OS

XX Streptococcus pneumoniae.

XX

PN WO200039299-A2.

XX

PD 06-JUL-2000.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 06:41:29 ; Search time 28 Seconds
(without alignments)
2734.588 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTVKENRRVSY.....KLLALLKGSNPFSSVSKKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : PIR_78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4163	100.0	802	E95136	conserved domain p
2	4159	99.9	828	E98004	hypothetical prote
3	3224	77.4	855	D98004	histidine Motif-Co
4	2797.5	67.2	819	E95136	conserved domain p
5	2772	66.6	839	E95115	conserved hypothet
6	2733	65.6	853	C97985	hypothetical prote
7	1246	29.9	1039	H95115	conserved hypothet
8	1243	29.8	1039	D97985	hypothetical prote
9	934	22.4	822	T46758	hypothetical prote
10	243	5.8	182	E97985	hypothetical prote
11	192.5	4.6	1390	T44004	hypothetical prote
12	176.5	4.2	1271	A45555	hypothetical prote
13	172	4.1	1241	D90011	trfA protein - sli
14	171	4.1	1233	S56271	Glutamate rich pro
15	168.5	4.0	1043	D84900	FntB protein [limpo
16	165	4.0	891	D84900	hypothetical prote
17	161	3.9	1420	A44361	hypothetical prote
18	159	3.8	2810	T22238	hypothetical prote
19	158	3.8	1650	T18444	hypothetical prote
20	157	3.8	910	S73361	dnaJ homolog prote
21	157	3.8	1658	T22238	hypothetical prote
22	157	3.8	2195	S61103	hypothetical prote
23	156.5	3.8	1495	T48429	SEC16 protein - ye
24	156	3.7	519	G84598	hypothetical prote
25	154.5	3.7	1338	T18416	probable bZIP tran
26	154.5	3.7	1463	T30290	hypothetical prote
27	154.5	3.7	1516	E71619	AAS surface protei
28	154.5	3.7	3924	S37431	RAD2 endonuclease
29	154	3.7	749	S23467	ankyrin 2, neurona
					probable long-chai

RESULT 1

C95136

conserved domain protein SP1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C/Accession: C95136
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide

ALIGNMENTS

30	151.5	3.6	1163	2	A36685	205K microtubule-a
31	150.5	3.6	1176	2	S66771	hypothetical prote
32	150	3.6	1132	2	B82538	ribonuclease E XF2
33	150	3.6	2401	2	T28676	phoptry protein -
34	149.5	3.6	1175	2	S51005	protein-tyrosine-p
35	149	3.6	1274	2	A89959	hypothetical prote
36	149	3.6	2829	2	A42771	reticulocyte-bindl
37	149	3.6	4688	2	B2885	hypothetical prote
38	149	3.6	6713	2	B89921	hypothetical prote
39	148.5	3.6	1785	2	T21558	hypothetical prote
40	148.5	3.6	1959	2	AG1085	hypothetical prote
41	148.5	3.6	1969	2	T38495	hypothetical prote
42	148	3.6	980	2	E71606	hypothetical prote
43	148	3.6	1230	2	T22458	hypothetical prote
44	148	3.6	1856	2	C95008	immunoglobulin A1
45	147	3.5	700	2	A56976	transfer complex p
46	147	3.5	1269	2	F84730	probable myosin he
47	147	3.5	1929	2	T21559	hypothetical prote
48	146.5	3.5	2004	2	F55133	immunoglobulin A1
49	146	3.5	821	2	S67087	hypothetical prote
50	145.5	3.5	1073	2	S14032	kinesin-related pr
51	145.5	3.5	1085	2	T38378	kinesin-like prote
52	145.5	3.5	1240	2	S52734	hypothetical prote
53	144.5	3.5	934	2	T47546	protein kinase-lik
54	144.5	3.5	4152	2	T31102	filamentous hemagg
55	144.5	3.5	4919	2	T31105	hypothetical prote
56	143.5	3.4	900	2	C64232	alanine-tRNA ligas
57	142.5	3.4	875	2	S70115	ZIP1 protein - yea
58	142.5	3.4	1066	2	T45283	growth polarity ma
59	142.5	3.4	1066	2	T41099	staurosporine targ
60	142.5	3.4	1979	2	C71622	hypothetical prote
61	142.5	3.4	5170	2	T15348	hypothetical prote
62	142	3.4	1257	2	T00486	serine/threonine-s
63	142	3.4	1744	2	JH0720	tanabin - African
64	142	3.4	1822	2	S33441	EF protein - Strep
65	141.5	3.4	490	2	S52830	HMS1 protein - yea
66	141.5	3.4	1315	2	T28679	fibrinogen-binding
67	141.5	3.4	1928	2	S46773	myosin heavy chain
68	141.5	3.4	2314	1	A46151	protein-tyrosine-p
69	141.5	3.4	2748	2	S57976	nuclear migration
70	141.5	3.4	3498	2	T22330	hypothetical prote
71	141	3.4	914	2	B48086	translation initia
72	141	3.4	1190	2	S47536	oxysterol-binding
73	141	3.4	1280	2	T42514	kinase anchor prot
74	141	3.4	1332	2	S41552	probable transcrip
75	141	3.4	2288	2	T29999	hypothetical prote
76	141	3.4	3890	2	C89921	hypothetical prote
77	140.5	3.4	507	2	S05542	hypothetical prote
78	140.5	3.4	719	2	A81358	hypothetical prote
79	140.5	3.4	2231	2	D71870	hypothetical prote
80	139.5	3.3	1535	2	T49042	hypothetical prote
81	139.5	3.3	1609	1	MMHUB2	laminin gamma-1 ch
82	139.5	3.3	4549	2	T20771	hypothetical prote
83	139.5	3.3	4667	2	T20774	hypothetical prote
84	139	3.3	1247	2	C89583	protein K07E3.1 [1
85	139	3.3	1385	2	D89824	hypothetical prote
86	139	3.3	1957	2	T38077	hypothetical prote
87	138.5	3.3	1703	2	S15047	SNF2 protein - yea
88	138.5	3.3	2269	2	T28677	thoptry protein -
89	138	3.3	1639	2	S05603	major merozoite su
90	138	3.3	2295	2	B71621	probable membrane

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: C95136
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-802 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:gl4972655; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: Spil75
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 100.0%; Score 4163; DB 2; Length 902;
 Best Local Similarity 99.9%; Pred. No. 1.7e-219;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNNRVSYIDGKQATQKTENLTDEVSKREGINAEQIVIKITDQGYVT 60
 DB 7 SYELGLYQARTVKNNRVSYIDGKQATQKTENLTDEVSKREGINAEQIVIKITDQGYVT 66

QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNYKDKEDI VNEVGKGYVVKDGYVYVLKDA 120
 DB 67 SHGDHYHYNGKVPYDAIISEELLMKDPNYKDKEDI VNEVGKGYVVKDGYVYVLKDA 126

QY 121 AHADNVRTKEENRQKQHSQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
 DB 127 AHADNVRTKEENRQKQHSQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 186

QY 181 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240
 DB 187 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 246

QY 241 VSNPGTTNTNNSNTNSQASQNSDIDSLKQYKPLSQRHVESDGLVDFDPAQITSTRT 300
 DB 247 VSNPGTTNTNNSNTNSQASQNSDIDSLKQYKPLSQRHVESDGLVDFDPAQITSTRT 306

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPDSRPEQSPQPTPEPSPG 360
 DB 307 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPDSRPEQSPQPTPEPSPG 366

QY 361 PQAPNKLKIDSNSLSVQLVRKVGEGVYFEEKISRYVFAKOLPSETVKNLESKLSQES 420
 DB 367 PQAPNKLKIDSNSLSVQLVRKVGEGVYFEEKISRYVFAKOLPSETVKNLESKLSQES 426

QY 421 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 DB 427 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 486

QY 481 KEKLVDDLLAFLAPITHPERLKGPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 540
 DB 487 KEKLVDDLLAFLAPITHPERLKGPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 546

QY 541 EGDAYVTPHMGSHHWIGKDSLSKDKVAAQAYTKKGIILPPSPDADVKANPTGDSAAAY 600
 DB 547 EGDAYVTPHMGSHHWIGKDSLSKDKVAAQAYTKKGIILPPSPDADVKANPTGDSAAAY 606

QY 601 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNKFAWFDHDTYKAPNGYTTLED 660
 DB 607 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNKFAWFDHDTYKAPNGYTTLED 666

QY 661 LFATIKYVEHPDRPHNSDGNWSEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
 DB 667 LFATIKYVEHPDRPHNSDGNWSEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 726

QY 721 QVETEKVEAQKAEAVLLAKVTDSLSKANATETLAGLNLLTLQIMDNNSIMAEKLLA 780
 DB 727 QVETEKVEAQKAEAVLLAKVTDSLSKANATETLAGLNLLTLQIMDNNSIMAEKLLA 786

QY 781 LLKGSNFPSSVSKKIN 796

Db 787 LLKGSNFPSSVSKKIN 802

RESULT 2

E98004

hypothetical protein phta [imported] - *Streptococcus pneumoniae* (strain R6)

C:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C:Accession: E98004

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burtett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98004

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-828 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:gl5458683; GSPDB:GN00174

C:Genetics:

A:Gene: phta

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 99.9%; Score 4159; DB 2; Length 828;

Best Local Similarity 99.7%; Pred. No. 2.9e-219;

Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNNRVSYIDGKQATQKTENLTDEVSKREGINAEQIVIKITDQGYVT 60

DB 33 SYELGLYQARTVKNNRVSYIDGKQATQKTENLTDEVSKREGINAEQIVIKITDQGYVT 92

QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNYKDKEDI VNEVGKGYVVKDGYVYVLKDA 120

DB 93 SHGDHYHYNGKVPYDAIISEELLMKDPNYKDKEDI VNEVGKGYVVKDGYVYVLKDA 152

QY 121 AHADNVRTKEENRQKQHSQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180

DB 153 AHADNVRTKEENRQKQHSQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 212

QY 181 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240

DB 213 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 272

QY 241 VSNPGTTNTNNSNTNSQASQNSDIDSLKQYKPLSQRHVESDGLVDFDPAQITSTRT 300

DB 273 VSNPGTTNTNNSNTNSQASQNSDIDSLKQYKPLSQRHVESDGLVDFDPAQITSTRT 332

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPDSRPEQSPQPTPEPSPG 360

DB 333 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPDSRPEQSPQPTPEPSPG 392

QY 361 PQAPNKLKIDSNSLSVQLVRKVGEGVYFEEKISRYVFAKOLPSETVKNLESKLSQES 420

DB 393 PQAPNKLKIDSNSLSVQLVRKVGEGVYFEEKISRYVFAKOLPSETVKNLESKLSQES 452

QY 421 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480

DB 453 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 512

QY 481 KEKLVDDLLAFLAPITHPERLKGPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 540

DB 513 KEKLVDDLLAFLAPITHPERLKGPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 572

QY 541 EGDAYVTPHMGSHHWIGKDSLSKDKVAAQAYTKKGIILPPSPDADVKANPTGDSAAAY 600

DB 573 EGDAYVTPHMGSHHWIGKDSLSKDKVAAQAYTKKGIILPPSPDADVKANPTGDSAAAY 632

QY 601 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNKFAWFDHDTYKAPNGYTTLED 660

DB 633 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNKFAWFDHDTYKAPNGYTTLED 692

QY	661	LFATIKYVVEHPDRPHSNDGWNASEHVLGKKHSDENPKNFKKADEEPEETPAEPEVP	720
Db	693	LFATIKYVVEHPDRPHSNDGWNASEHVLGKKHSDENPKNFKKADEEPEETPAEPEVP	752
QY	721	QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTLOIMDNNSINAAEAKLLA	780
Db	753	QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTLOIMDNNSINAAEAKLLA	812
QY	781	LLKGSNPSSVSKEKIN	796
Db	813	LLKGSNPSSVSKEKIN	828
RESULT 3			
D98004			
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)			
C:Species: Streptococcus pneumoniae			
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001			
C:Accession: D98004			
E:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E			
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M			
e, P.; Sun, P.M.; Winkler, M.E.			
J. Bacteriol. 183, 5709-5717, 2001			
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;			
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.			
A:Reference number: A97872; MUID:21429245; PMID:11544234			
A:Accession: D98004			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-855 <KUR>			
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:gl5458682; GSPDB:GN00174			
C:Genetics:			
A:Gene: phpA			
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein			
Query Match 77.4%; Score 3224; DB 2: Length 855;			
Best Local Similarity 75.0%; Pred. No. 3.2e-168;			
Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;			
QY	1	SYELGLVQARTVKENNVSVYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT	60
Db	32	AYELGLHQATVKENNVSVYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT	91
QY	61	SHGDHYHYNGKVPYDAIISELLMKDPNPKLDEIVNEVGKGYVTKDGKYYVYLKDA	120
Db	92	SHGDHYHYNGKVPYDAIISELLMKDPNPKLDEIISEIKGKGYVTKDGKYYVYLKDA	151
QY	121	AHADNVRTKEINRQKQHSQHREGGTPNDGAVALARSGQRYTTDDGYIFNASDIIEDT	180
Db	152	AHADNVRTKEINRQKQHSQHREGGTPNDGAVALARSGQRYTTDDGYIFNASDIIEDT	211
QY	181	GDVIVPHGDHYHYIPKNELSAELAAEAFISGRGNLSNRYRRQNSDNTSRTNWVPS	240
Db	212	GDVIVPHGDHYHYIPKNELSAELAAAFISGRGNLSNRYRRQNSDNTSRTNWVPS	271
QY	241	VSPNGTTNTNTSNNSTNSQASQSDINDSLKQLYKPLSQRHVESDGLVFPDAQITSR	300
Db	272	VSPNGTTNTNTSNNSTNSQASQSDINDSLKQLYKPLSQRHVESDGLVFPDAQITSR	331
QY	301	ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQSTPEPSPG	360
Db	332	ANGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQSTPEPSPG	391
QY	361	POPAPNLK-IDSN--SSLVSQLVRKVGEGYFKEGTSRYVPFADLPSETVKNLESKLS	416
Db	392	POPAPNPQAPSNPIDSKLVKEAVKVGEGYFKEGTSRYVPFADLPSETVKNLESKLS	451
QY	417	KQESVSHTLTAKENVAFRQDFYDKAYNLLTEAHKALFXNKGNSDFCALDKLLERLND	476
Db	452	KQESLSHLGAKXTDLPSSDREFYNKADLLARIHQDILLDNKGRQVDFEALONLLERLND	511
QY	477	ESTNKEKLVDLLAFLAPIITHELKGNPSQIETDEVEVIAQLADKYTTSDGYTFDEHD	536

Db 306 TARGVAVPHGNHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPDRPEPSPQTPPEPSP 365
 QY 360 GPAPAPNLKIDSSNVLSQVRKVGEGYVFEKGIERYVPAKDLPSBTVKNLSEKSKOE 419
 Db 366 SPOPAISNPD--EKLKAVKRVGVDGYFEENGVSRYIPAKDLSAETAAGIDSKLAKOE 423
 QY 420 SVSHTLTAKENAVAPRQDEYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDEST 479
 Db 424 SLSHKLTGTTKTDLPSSDRFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLEKLDVSS 483
 QY 480 NKEKLVDDLLAFIAPITHPERLCKNSQIETDEVRIAQLADKYTSDGYIFDEHDIIS 539
 Db 484 DKVKLVEDIIAFIAPIRHPERLCKPNAQIYTTDEIQVAKLAGKYTTEDGYIFDPDITS 543
 QY 540 DEGDAYVTPHMGSHWTKDLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAAI 599
 Db 544 DEGDAYVTPHMTSHWIKDLSAEARAAQAAYAKEKGLTPSTDHQDSGNTEAKGAEAI 603
 QY 600 YNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAWFDHTYKAPNGYTLLE 659
 Db 604 YNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWDEGLYEAPKGYILE 663
 QY 660 DLFTATIKYVVEHPDERPHSDNGMGNASEHVLKGDHSEDPMKPKADEE-----PVEET 713
 Db 664 DLLATVYKYVVEHPNERPHSDNGMGNASDHVORNGNQADTNQTEKPSSEKPKQTEKPREET 723
 QY 714 -----PAPEVPQVETEKVBAQIKRAEVILAKVTDSSLKANA 750
 Db 724 PREKPOSEKPEPKTEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 783
 QY 751 TETLAGRNUTLQIMDNNSIMAEAKLLALLKGS 785
 Db 784 KETLTGLKNLLFGTDNNNTIMAEAKLLALLKGS 818

RESULT 5
 G95115
 conserved hypothetical protein sp1003 [imported] - Streptococcus pneumoniae (strain TIGR
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
 neon, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; PMID:21357209; PMID:11463916
 A:Accession: G95115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:g14972476; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: Sp1003
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 66.6%; Score 2772; DB 2; Length 839;
 Best Local Similarity 65.0%; Pred. No. 1.4e-143;
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATKTNLTTPDEVSKREGINAEQIVIKITDQGV 59
 Db 21 SYELGRHQAGQVKESNRVSYIDGQAGKAENLTTPDEVSKREGINAEQIVIKITDQGV 80
 QY 60 TSHGDDHYNGKVPYDAIISBELLMKDPNYKLKDDIYNEVKGGVYKVDGKYVYLLKD 119
 Db 81 TSHGDDHYNGKVPYDAIISBELLMKDPNYQLKSDIYNEIKGGYVYKVDGKYVYLLKD 140
 QY 120 AAHADNVRTKEEINRQKQESQHRREGTTPNDGAVALARQGRYTTDDGYIFNADIIED 179
 Db 141 AAHADNVRTKEEIKRQKQESHNHGGGS--NDQAVVAARAQGRYTTDDGYIFNADIIED 198

QY 180 TGDAYIVPHGDHYHYIIPKNELSASELAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVP 239
 Db 199 TGDAYIVPHGDHYHYIIPKNELSASELAAAEAFYNG-----KQSRPESSSSSYNA 247
 QY 240 SVSNPTGTTNTSNNNTNSQASQSDNDISLLKQLYKLPLOSORHVESDGLVDFPAQITSR 299
 Db 248 NPAQPRLSNNHLLTVPTTYHQ-NOGENISLLRELYAKPLSERHVSDDGLIFDPAQITSR 306
 QY 300 TARGVAVPHGDHYHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPDRPEPSPQTPPEPSP 359
 Db 307 TARGVAVPHGNHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPDRPEPSPQTPPEPSP 366
 QY 360 GPAPAPNLK-IDSN-----SSLVSQLVRKVGEGYVFEKGIERYVPAKDLPSBTVKNLSEK 415
 Db 367 SPOPAISNPD--EKLKAVKRVGVDGYFEENGVSRYIPAKDLSAETAAGIDSKL 426
 QY 416 SKQESVSHUTITAKENAVAPRQDEYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLN 475
 Db 427 AKQESLSHLKLGAKKTDLPSSDRFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLEK 486
 QY 476 DESNKEKLVDDLLAFIAPITHPERLCKPNSQIETDEVRIAQLADKYTSDGYIFDEH 535
 Db 487 DVPSDKVKLVDDLLAFIAPIRHPERLCKPNAQIYTTDEIQVAKLAGKYTTEDGYIFDPR 546
 QY 536 DIISDEGDAYVTPHMGSHWIKDLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDS 595
 Db 547 DIISDEGDAYVTPHMTSHWIKDLSAEARAAQAAYAKEKGLTPSTDHQDSGNTEAKG 606
 QY 596 AAATYNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAWFDHTYKAPNG 655
 Db 607 AEATYNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWDEGLYEAPK 666
 QY 656 YTLRDLFATIKYVVEHPDERPHSDNGMGNASEHVLK-----KQHS----- 697
 Db 667 YTLRDLFATIKYVVEHPNERPHSDNGMGNASDHVRKKNQVQDQSKPDEDKEHDEYSEPTHP 726
 QY 698 -----DNKPKADEPEVEETPAPEVPQVETEKVBAQIKRAEVILAKVTDSSLKANA 738
 Db 727 ESDEKENHAGLNPSADNLKPSDTEETEEAEADTTDEAIPQVENSVINAKIADREALL 786
 QY 739 AKYVTDSSLKANATETLAGRNUTLQIMDNNSIMAEAKLLALLKGSNPSP 790
 Db 787 EKVTDPISRONAMETLTGLKSSLLGTGKDNNTIISAEVDSLLALLKESQFAP 838

RESULT 6
 C97985
 hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: C97985
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; Dehoff, B.S.; E
 e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; PMID:21429245; PMID:11544234
 A:Accession: C97985
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-853 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:g15458514; GSPDB:GN00174
 C:Genetics:
 A:Gene: phtD
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 65.6%; Score 2733; DB 2; Length 853;
 Best Local Similarity 63.5%; Pred. No. 2e-141;
 Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATKTNLTTPDEVSKREGINAEQIVIKITDQGV 59

Db	21	SYELGRHQAGVQKYESNRVSVIDGDAQKAENLTPDEVSKREGINAEQIVIKITDQGYV	80
Qy	60	TSHGDGHVHYNGKVPYDAI ISEELLMKDPNYKLKDEDIVNEVKGGVIVKDGKYYVYLKD	119
Db	81	TSHGDGHVHYNGKVPYDAI ISEELLMKDPNYQLQSDSDIVNEIKGGVIVKDGKYYVYLKD	140
Qy	120	AAHADNVRTKEEINRQKQEHSHQREGGTFRNDGAVALARSOGRYTTDDGYIIFNADSLIT	179
Db	141	AAHADNIRTKEEIKRQKQERSHNHS ---RADNAVAARAQGRYTTDDGYIIFNADSLIT	197
Qy	180	TGDAIVTPHGDGHYIIPKVELSASELAAEAAPLSGRGNLSNRTYRRQNSDSTRTNWVP	239
Db	198	TGDAIVTPHGDGHYIIPKSDLSASELAAQAQYWG- -----KQGRSPRSSSSHNA	246
Qy	240	SVSNPGTNTNTSNNSNTNSQASOSNDIDSLKLQYKLPLSQRHVESDGLVPDPAQITSR	299
Db	247	NPAQPLRSEHNLTVTYTYIQ -NOGENISSULRELYAKPLSERHVESDGLIFDPAQITSR	305
Qy	300	TARGVAVPHGDGHYHFPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSP	359
Db	306	TANGVAVPHGDGHYHFPYSQLSPLEEKLARIIPLYRSNHWVPSRPEQSPQPTPEPSP	365
Qy	360	GPQAPNLK-IDSN---SSLVSQVLRKVGEGYVFEKGISRYVFAKOLPSETVKNLESKL	415
Db	366	SQPAENPOAPENPIDEKLVKAVRKGVDGYVFEENGVPRIYIPAKOLSAETAAGIDSKL	425
Qy	416	SKOESYSHLTAKENVAVPDQEFYDKAYNLLTEAHKALFANKGRNSDFQALDKLLERLN	475
Db	426	AQGESLSHLGAKKTDLPSSDRFPYKAYDILLARIHQDLDLNKGRQVDFEALDNLLERLK	485
Qy	476	DESTNKEKLVDDLAFAPITHPERLGKPNSEQIYTEYDEVRIAQLADKYTTSDGYIFDEH	535
Db	486	DVSSDKVKLVDDILAFAPIRHPERLGKPNQIYTTDDEIQVAKLACKYTTEDGYIFDPR	545
Qy	536	DIISDEGDYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKGLIPPSFDADVKANPTGDS	595
Db	546	DITSDEGDYVTPHMTSHWIKDLSSEAEAAAAQAYAKEKGLTPPSTDHQDSGNTBAKG	605
Qy	596	AAAIYNRVKGEXRIPLVRLPYMVEHVVEKNGNLIIPHKHYNINIKFAPWDDHTYKAPNG	655
Db	606	AEAIYNRVKAAKVPLDRMPYQLQYVVEKNGSLIIPHVDYHNIKFEPWDEGLYEAPKG	665
Qy	656	YTLDELFIATIKYVVEHPDRPHSDNCGWNASAEHVLGK- -----DHSEDPNK- -----	701
Db	666	YSLDLLATVKYVVEHPDRPHSDNCGFNASDHVQENKNGQADTNTQTEKPNEEKPQTEKP	725
Qy	702	-----NFKAD-----EEPEETPAPEVPEQVE	723
Db	726	EEDKEHDEVSETHPDESDEKENHVGLNPSADNLKYPSTDTEETEEBAEDTTDEAEIPQVE	785
Qy	724	TEKVQAQLKEAEVLIAKVTDSLSKANATETLAGLRNLTLOIMDNNSIWAEBEKLIALLK	783
Db	786	HSVINAKEAEAEALLEKVTDSIRQNAVETTLGLKSSLLGTGDNNNTISAEVDSLIALLK	845
Qy	784	GSNPSSV 790	
Db	846	ESQPTPI 852	

RESULT 7
H95115
conserved hypothetical protein Spi004 [imported] - Streptococcus pneumoniae (strain TIGR
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; PMID:21357209; PMID:11463916

A;Accession: H95115
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1039
A;Cross-references:
A;Experimental source:
C;Genetics:
A;Gene: SP1004

Query Match 29.9%; Score 1246; DB 2; Length 1039;
Best Local Similarity 40.9%; Pred. No. 3.6e-60;
Matches 296; Conservative 97; Mismatches 180; Indels 150

Qy	1	SYELGYQARTYKNNRRYSYIDGQATOKTENLTPDEVSKEGREGNABQIVIKITDQGYVT	60
Db	22	AVALLQHRSQENKNNRRYSYVGGSSQSKSENLTFDQVSQKQIQABQIVIKITDQGYVT	81
Qy	61	SHGDHYHYNGKVPYDALIIBELMKDNYKDKEDIVNEVKGYIVKDVGYVYLKDA	120
Db	82	SHGDHYHYNGKVPYDALFSELMLKDPNYQLKDADIVNEVKGYIIVKDVGYVYLKDA	141
Qy	121	AHADNVRTKEELNRQKQSHSQHREGTTPRNDGAVALARSOGRYTTDDGYIFNADIIEDT	180
Db	142	AHADNVRTKDEINRQKQSHVADNE-----KVNNSVAVARSQRYTTDNDGYVFNPAIIEDT	197
Qy	181	GDAYIVPHGDHYHYIPKNEISASELAAEAFLSGRGLNSRTRYRQNSDNTSRTNWVPS	240
Db	198	GNAYIVPHGHHYVYIPKSDLSASELAAKAHLAGKNQPSQLSYSTASDN-----	248
Qy	241	VSNPCTTNTNSNSNTNSQASQNDIDSLIKOLYKLPISQRHVESDGLVFPQAQITST	300
Db	249	-----NTQSVAKGSGTSPANKSENLOSLKELYDSPSAQRYSESDGLVFPQAQITST	301
Qy	301	ARGVAVPHGDHYHPIPYSSOMSELEERTARILPIRYRSNHWVDSRPEQSPQTPPEPSG	360
Db	302	PNGVAIPHGDHYHPIPYKLSALDEEKIARMVP-----	333
Qy	361	POPAPNIKIDNSLSLVSQLVRKVGEGYVFEKGISRYVFAKDLPSETVKNLKSLKQES	420
Db	334	-----ISGTGSTVSTNAK-----PNEVSSIGLSLSNPS	363
Qy	421	VSHTLTAKKENVAPROQEFYDKAYNLLTBAHKALFANKGRNSDFQALDKLLERLNDESTN	480
Db	364	-----LITSKELSSASDGYIFNPK-DIVEETAAYIVRHG--DHFHYIPK-----SNQIG	410
Qy	481	KEKLVDDLLAFAPITHPERLKGPNISQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD	540
Db	411	QPTLPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDANRIIAE	452
Qy	541	EGDAYVTPMGHSHWIGKDSLSDKEKVAQAAYTKEKGIPLPPSDADVKANFTGDSAAIY	600
Db	453	DESGFVMSHGHDHNYFFKDXLTTEEQIKAAQKHLE-----VTSINGLDSLSH	501
Qy	601	NR-----VKGEKRIPLVRLPMV--EHTVEVKGNLIJ-PhKHdYHNiKfAW	644
Db	502	EQDYPNAKEMKDLKKIEEKIAGIMQYGVKRESIVVNEKNAIYPHGDHHPADP----	558
Qy	645	FDHTYKAPNGYTLBEDLPATIIYVYEHDPDRPHSNDG-----WGNASEHVLKQKHSBDPN	700
Db	559	IDEH---KPVGIG---HSHSNYELFKPEEGVAKKGNKVYTGBELTNVNLKNSTFNN	611
Qy	701	KNP 703	
Db	612	QNF 614	

RESULT 8

D97985
hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; De

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AB007317; PIDN:AAK99712.1; PID:G15458515; GSPDB:GN00174
C:Genetics:
A:Gene: pntE

Query Match 29.8%; Score 1243; DB 2; Length 1039;
Best Local Similarity 41.2%; Pred. No. 5.2e-60;
Matches 298; Conservative 94; Mismatches 180; Indels 152; Gaps 20;
3Y 1 SYELGLYQARTVKNRVSVIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
3D 22 AYALNQRHSQENKDNRRVSVDGSSQSKGENLTPDQVQKGLQAEQIVIKITDQGYVT 81
3Y 61 SHGDHYHYNGKVPYDAIIESEELMKDPNKLKDEDIVNEVKGGYVIKVDGKYVYLKDA 120
3D 82 SHGDHYHYNGKVPYDALFSEELMKDPNQLKQADIVNEVKGGYVIKVDGKYVYLKDA 141
3Y 121 AHADNVRTKEINRQKQEHSGHRRGGTTRNDGAVALARSGRYTTDQGYIFNADITIEDT 180
3D 142 AHADNVRTKEINRQKQEHVKNDE---KVNNSVAVARSQRYTTDQGYFNPADITIEDT 197
3Y 181 GDVIVPHGDHYHYIPKNELASASELAEEFLSGRNLNSRTVRRQNSDNTSRTNWVPS 240
3D 198 GNAYIVPHGGHYHYIPKSDLSASELAEEFLSGRNLNSRTVRRQNSDNTSRTNWVPS 248
3Y 241 VSNPGTNTNNSNTSQAQSDNDISLKLQYKLPFLSRHVESGLVFDPAQITSR 300
3D 249 ---NQVAKGSTKPKANKSENLOSLLKELYDSPSAQRYSESDGLVFDPAQITSR 301
3Y 301 ARGVAVPHGDHYHYIPYSQSELEERARIIPLYRSHWVDSRPEQSPQTPPEPSG 360
3D 302 PNGVAIPHGDHYHYIPYKLSALEEKIARRVP-----PNEVSSLSLSSNPS 363
3Y 361 POPAPNLKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSQES 420
3D 334 ---ISGTGSTVSTNAK-----PNEVSSLSLSSNPS 363
3Y 421 VSHLTAKENVAPRQDFYDKAYNLLTEAHKALFKNKGRNSDFQALDKLLERLNDESTN 480
3D 364 ---LTSKELSSASDGYIFNPK-DIVETATAYIVRHG--DHFHYVTPK-----SNQIG 410
3Y 481 KEKLVDDLLAFIAPITHPERLQKP-NSQIEYTEDVRIQLADKYTTSDGYIFDEHDIIS 539
3D 411 QPTLPNNSLATPSP-----SLPINPGISHEKHE-----EDGYGFANDRIIA 451
3Y 540 DEGDAYVPHMGSHWIGKDSLDKQVAAQAYTEKEGILPPSPDADVKANPTG-DSAAA 598
3D 452 EDESGFIMSHGNHNYFFKDLTEEQIKAAQKHLE-----VKTSHGLDLSLS 500
3Y 599 IYNRVKG-----EKRIPLVRLPYM--EHTVEKGNLII-PKHQHYHNKIFA 643
3D 501 HEQDYPGNKEMKDLKKEIKIAGIMQYGVKRESIVVNEKNAIIPYHGDHHDADP-- 558
3Y 644 WFDHTYKAPNGYTLDELATIKYVYEHDPDRPHSNDG---WGNASEHVLGKKHSDP 699
3D 559 -IDEH---RPVIGG---HSHSNYELFKPEGVAKKEGNKVYTGEBLTNVNLLKNSTFN 610
3Y 700 NKNF 703
3D 611 NQNF 614

hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Spellerberg, B.; Rodzinski, E.; Martin, S.; Weber-Heynenmann, J.; Schnitzler, N.; Luett
Infect. Immun. 67, 871-878, 1999
A:Title: Lmb, a protein with similarities to the Lral adhesin family, mediates attachment
A:Reference number: Z24091; MUID:99115568; PMID:9916102
A:Accession: T46758
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPE>
A:Cross-references: EMBL:AF062533; NID:G4249622; PIDN:AAD13797.1; PID:G4249624
A:Experimental source: strain R268
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.4%; Score 934; DB 2; Length 822;
Best Local Similarity 29.4%; Pred. No. 2.7e-43;
Matches 271; Conservative 99; Mismatches 235; Indels 318; Gaps 32;
3Y 1 SYELGLYQARTVKNRVSVID---GKQATQKTENLTPDEVSKREGINAEQIVIKITDQ 57
3D 22 SYQLGKXHMGLATKDNQIAVIDDSKGVKAPKT-NKTMQISAEEGISAEQIVVKITDQ 80
3Y 58 YVTSHGHDHYHYNGKVPYDAIIESEELMKDPNKLKDEDIVNEVKGGYVIKVDGKYVYL 117
3D 81 YVTSHGHDHYHYNGKVPYDAIIESEELMTDPNHFQKSDVINELDGYVIKVGNGYVYL 140
3Y 118 KDAHADNVRTKEINRQKQEHSGH-REGG-----TPRNDGAVALARSGRYTTDQGY 169
3D 141 KPGSKRNKIRTKQIAEQVAKGKEKEGLAQVAHLKSEEAANVNEAKRQGRYTTDQGY 200
3Y 170 IFNASDIETDGDVIVPHGDHYHYIPKNELASASELAEEFLS---GRGNLSNRTVRR 226
3D 201 IFSPTDIIIDDLGDAYLPHGNHYHYIPKDLSPSELAAQAYWSQKQGRG--ARPSDVRP 258
3Y 227 QNSDNTSRTNWVPSV-NPGTNTNTN-NSNTNSQAQSDN-----IDSLK 272
3D 259 TPAFGRKAP-IPDVTNPGQGHQPDNGHYHAPPRNDASQNKHQDEFKGTFKELLD 317
3Y 273 QLYKLPISQRHVESDGLVFDPAQITSRGAVPHGDHYHYIPYSQSELEERARIIP 332
3D 318 QLRLDLKYRHEDEGLIFETQVKSNAFGVPHGDHYHYIPRSLSPLEMLAD--- 374
3Y 333 LYRSNHWVDSRPEQSPQTPPEPSGPQAPNLKIDNSLSVQLVRKVGEGYVFEK 392
3D 375 -RYLAG-----QTDNDN-----GSDH----- 390
3Y 393 GISRVVFAKDLPSRTVKNLESKLSQESVSHLTAKENVAPRQDFYDKAYNLLTEAHK 452
3D 391 -----SKPSDK- VTHFTLGR-----IKAY----- 410
3Y 453 ALFPXNKRNSDFQALDKLLERLNDESTNKEKLVDDLLAFIAPITHPERLGRNSQIEYTE 512
3D 411 -----GKGLD-----GK----- 418
3Y 513 DEVRIAQLADKYTTSDGYIFDEHDIISDEGDVAVTPHMGSHWIGKDSLDKQVAAQAY 572
3D 419 -----YDTSDAYVFSKESHSVDKSGVTAKGDDHFFHYIGFGELEQVLEDEVANW 467
3Y 573 TKEG-----ILPPSPDADVKANPTGDSAAA-----TYNR----- 602
3D 468 VKAQGADELVAALDOEQKEKELFDTKVKSRKVTGDKGVYIMPRDKDYFYARQLDL 527
3Y 603 -----VKGEKR-----IPLVRLPYMVEHTVEKGNLIIIPKH 636
3D 528 TQIAFAEQELMLKDKKHRYDYIVDTGIEPLRAVDVSSPLMHAGNATYDTGSSFVPHIDH 587
3Y 637 YHNKFAWFDHTYKAPNGYTLDELATIKYVYEHDPDRPHSNDG-----GNASEH 688
3D 588 IHVVPYSWL-----TNQIATIKYVMOHVEVP--DVMSKPGHBSGSGVIFN 632
3Y 689 VLKCKHSDPEDPNKNFKADEEPVETPAE-----PE-----VP 720

Db 633 VTPLDKRGMENWQIIHSAEVEQKALAEGRFAAPDGYIFDPRDLAKETFFWKDGSFSP 692
 QY 721 QVE-----TEKVEAQLKEAEVLLAKVTDSLLKANATETLAGRNLTILQIMDNNSI 771
 Db 693 RADGSSLTINKSDLSQAEWQAQELLAKN-----AGDATDT-----DKPEE 735
 QY 772 MAEAEKLIALKGNPSSVSKK 794
 Db 736 KQADK---SNENQOPSSEASKEE 755

RESULT 10
 P97985
 hypothetical protein phxB-truncation [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: F97985
 R:Hoskins, J.A.; Albarn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:1154234
 A:Accession: F97985
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-182 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:gl5458517; GSPDB:GN00174
 C:Genetics:
 A:Gene: phxB-truncation

Query Match 5.8%; Score 243; DB 2; Length 182;
 Best Local Similarity 31.6%; Pred. No. 1.5e-06;
 Matches 61; Conservative 16; Mismatches 50; Indels 66; Gaps 5;
 QY 147 TPNNDGVALARSQRYTTDGGYIFNASDIIEDTGDAYIVPHGDHYHYPKNSLSASELA 206
 Db 4 TPNN-----GVSADVDDGVFNPNDIVRDGTDAYIVRHGDHYHYPKNSLIN----- 48
 QY 207 ABAEFLSGRGNLSRTRYRQNSDNTSRNTWVSVNPGTNTNTNNSNTNSQASQSDND 266
 Db 49 -----NPPSHSNTTEVSGSSSS 65
 QY 267 IDSLKQLYKPLSRHVESDGLVDFPAQITRTARGVAVPHGDHYHYPYSQMSLEER 326
 Db 66 V-----LNPSPHVVHBEEDGDFDANRIISDSRGFVIPHGDHNYIKV-QTKGYEAA 118
 QY 327 IARIIPLYRSNH 339
 Db 119 LKNKIP-SLQSNY 130

RESULT 11
 T14004
 trfA protein - slime mold (Dictyostelium discoideum)
 C:Species: Dictyostelium discoideum
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
 C:Accession: T14004
 R:Saito, J.; Adachi, H.; Sutoh, K.
 J. Biol. Chem. 273, 24654-24659, 1998
 A:Title: Dictyostelium TRFA homologous to yeast San6 is required for normal growth and e
 A:Reference number: Z17852; MUID:98406112; PMID:9733762
 A:Accession: T14004
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1390 <SAI>
 A:Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BA433143.1
 C:Genetics:
 A:Gene: trfA
 A:Introns: 333/3; 364/3; 637/1

Query Match 4.6%; Score 192.5; DB 2; Length 1390;
 Best Local Similarity 18.1%; Pred. No. 0.016;
 Matches 147; Conservative 131; Mismatches 291; Indels 245; Gaps 35;
 QY 15 NNRVSYIDGKATQKTENLTDPDEYSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVP 74
 Db 690 NNN 736
 QY 75 YDAIIEELMKDPNYLKDEDIVNEKGGYIKV-----DGKYYVYLKDAHAADNVRT 128
 Db 737 -----DNVSKNDVLDRLRYKGLEREKTSPPNGDR-----DREDNTRD 775
 QY 129 KEEINRQKQHSQHREGTTP-----NDGAVALA-----RSQGRYTTDDG 168
 Db 776 NRD-NRDSRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGR 834
 QY 169 YIFNASDIIEDTGDAYIVP---HGDHYHYPKNSLSASELAEEAFSLGRNLSNRYR 225
 Db 835 NNN 885
 QY 226 RQNSDNTSRNTWVSV-----SNPGTNTNTNNSNTNSQASQSDNDID-SLLKQL 274
 Db 886 NNN 941
 QY 275 YKLPLSORHVESDGLVDF-PAQITSR-ARGVAVPHGDHYHYPYSQMSLEERARIIP 332
 Db 942 ALSFQSQHKDRREIILDEESDINERSKTRPSI-----VKEAEKRETVIV 988
 QY 333 LRVSRNHWVDSRPEQSPQTPPEPSPQAPAPNLKI-DSNSSILVSVOLVRKVGEGYFEE 391
 Db 989 DKERS-----PTIITEKPEKQVEKTKESLVEKVDKE-----NEK 1027
 QY 392 KGISRYVFAKPLSETVKNLESKLSQESVSHLTITAKENVAAPRDOEFYDKAYNLLTEAH 451
 Db 1028 ESPSSSSSSSEIEKE 1078
 QY 452 KALFXNKGRRNSDFQALDKLLER-----LNDESTNKEKLVDLLAFIAPITHPERLGKN 505
 Db 1079 KDVEENK---SVEKSEKPEKESITTTTNDDEEGE-----LSEPTTTTKD 1123
 QY 506 SQEYETDEVRIAQLADKYITSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDE 565
 Db 1124 DSKLPTDEKLLSVSPVTTAVEQ-----SRDETKELE 1156
 QY 566 KVAQAQYTKEGILPPSPDADVKANPTGDSAAA-IYNRVKGEKRIPLVRLPYMVEHTVEV 624
 Db 1157 MDTKEDSEKEK-----KSSITTTAAASESVKPIDEEKKSPVTT-----TTTT 1198
 QY 625 KGNLILIP-HKDHYNHNIKFAWFDHDT-----YKAPNGYTLDELDFATIKYVVEHPD 673
 Db 1199 TMTTVEPTHKOKESKN-----DDTTTTTTTTTTKSAKSPNSPSTRS-----D 1242
 QY 674 E-RPHSNDGWNASEHVLGKDHSD-----PNKPFKADEEPPVEETPA-----EPVPOV 722
 Db 1243 EVVEPHQ-----DAQEEINKKLEDDITSTSKRLKPDSTFSSATTASTPSEQPESP-L 1296
 QY 723 ETEKVEAQLKEAEVLLAKVTDSLLKANATETLAG 756
 Db 1297 KKENPVGETLSLPEIKDKSSSSSSSSSSSSSTWTG 1330

RESULT 12
 A45555
 glutamate rich protein - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
 C:Accession: A45555; S27831
 R:Borre, M.B.; Dzlegiel, M.; Hogh, B.; Petersen, E.; Rieneck, K.; Riley, J.F.;
 J.
 Mol. Biochem. Parasitol. 49, 119-131, 1991
 A:Title: Primary structure and localization of a conserved immunogenic plasmodium falcipar
 rate life cycle.
 A:Reference number: A45555; MUID:92131041; PMID:1775153

A;Accession: A45555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1271 <BOR>
A;Cross-references: EMBL:M59706; NID:G160311; PID:G160312
A;Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBI:77802)

Query Match 4.2%; Score 176.5; DB 2; Length 1271; Best Local Similarity 19.1%; Pred. No. 0.11; Matches 170; Conservative 140; Mismatches 347; Indels 231; Gaps 43;

QY 7 QYATVKNRVSVIDGKQATKQKTENLTP-----DEVSKREGINAEQIVIKITDOGY---- 58
DB 373 HNINVLQNNINNHQLEPQEPKPNIESPEPKNIDSEILPENVEEIIIDVPSKHSNHE 432
QY 59 ----VTGSHGHHYNGKVPYDAIISELLMKDPNKLKOBIDIVNEVKGKGVVIVKDGKYY 114
DB 433 TFEETSESEHEEAVSEKNAHEITVEHEETVQESNPEKADND-----GNV----- 477

QY 115 VYLKDAHADN-VTKKEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNA 173
DB 478 ----SONSNELNENEFVESEKSEHE-----AAENEESLEEGLH---H 513

QY 174 SDII-----EDTGAYIVPHGDHVIYIPKNELSAELAAAEFLSGRNLNSRTYRQ- 227
DB 514 EEIPEQNNESGESKLV-----DNEGGEFAHNFSS---EVSNSLNENEF 560

QY 228 -NSDTSRTNWPVSNSFGTTNTNTSNNSNTNSQASQNDISLLKQYKLPQSRHVES 286
DB 561 VESDK-----SVTEPAHEEVSESNPEAENESSIEEAHOE--EIVPEQNDDES 610

QY 287 ---DGLV-----FDPQITIS-----RTARGVAVPHGDHVIYIPVSQMS 322
DB 611 GESGLVDNEEGDFEENHEEPDQNDSELSENLVESEKVSSEP-AEHVEIVSEKVSSE 669

QY 323 LEERTARIIPLYRNHNHWPDSRPPQPPQPTPEPS---PGPQAPNL-----KIDSNSL 375
DB 670 PAEHV-EIVSEKSTSE---PAHVESVSEQSNNESEKKDGPVPSKPEIEKVDVQPKI 725

QY 376 VSQVRKVGEGYVF-----EEKIGRYVPAKLPSETVK---MLESKLSQESVSHLTAK 428
DB 726 VD-LQIIEPNFVDSQNPQEPVPSFKIEKVPSEENKHASVDPEVKEKENVSEVEEK 783

QY 429 K-----ENVAPRDOFFYDKAYNLLTEAHK-----LFXNKGNSDFQALD 468
DB 784 QNSQESVEEIPVNEDEFEDVHTQDLQDHTVDPEIVEVEEIPSELHENEVAHPEIVEIE 843

QY 469 KLLERLNDESTNKEKLVDDLLAFLA-PITHPERLAKPNSQIERYTEDE-VRIAQLADKYTT 526
DB 844 EVFPENQNNNEQNEEDDKSAHIQHEIVEVEEILPEDDKNEKVEHEIVEVEEILPEDKN 903

QY 527 SDGYIFDEHDIISDEGDVYTPHMGSHWIGKDSLDEKVAQAAYTKEGILPPSPDAD 586
DB 904 EKG-----QHEIVEVE---EILPE-----DDKNEKVEHEIVEVEEILP--EDKN 942

QY 587 VKANPTGDSAAAIYNEVKGEKRIPLVPLPMVEH-TVEVKNGNLIIPHGDHVIYIKFAWF 645
DB 943 ERGOHEIVEVEEILPEDKNEK-----VEHEIVEVEE---ILPEDKNEKG----- 983

QY 646 DHTYKAPNGYTTLEDLFATIKYYVEHPDE-RPHSNDGWGNAS-----EHLVGKKDHE--- 697
DB 984 -QHEIVEVEEILPEDKNEKQVHEIVEVEEILPEDKNEKQHEIVEVEEILPEDKNEKGQ 1042

QY 698 -----DPNKPKADEEPEETPAPEVDQVTEKVEAQLKAEVLLAKVTDSLLK 747
DB 1043 HEIVEVEEILPEDKNEKQVHEIVE---VEEILPEDKNEKQVHEIVEVEEILPIVE----- 1095

QY 748 ANATEITLAGLRNLLTIQIMDNNSIMAEKLLALLKGSNPPSVSKSKI 795
DB 1096 --IEEVPFSQNNNENIE-----TIKPEKK-----NEFSVEEKAI 1128

D90011
FmtB protein [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: D90011
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: D90011
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2481 <KUR>
A;Cross-references: GB:BA000018; PID:G13701961; PIDN:BA843253.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: fmtB (mrp)

Query Match 4.1%; Score 172; DB 2; Length 2481; Best Local Similarity 19.4%; Pred. No. 0.5; Matches 169; Conservative 140; Mismatches 353; Indels 210; Gaps 38;

QY 8 QARTVKNRVSVIDGKQATKQKT-----ENLTPDEVSKREGINAEQIVIKITDOGYVTS 61
DB 969 QAATTKKSDAKAEIAQKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1018

QY 62 HGD-----HHVYNGKVPYDAIISELLMKDPNKLKDEIDIVNEVKGKGVVIVKDGKYY 115
DB 1019 NADINDTANTVDNAKTITNEATTA--AITPDANVKAQAIAD-----KVQAQ--- 1066

QY 116 YLKDAHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNA 159
DB 1067 --ETAIDANNSTTEEEAAKQVQTEKTA--DAIDAASHNVEVEAAKNABIAKIE 1120

QY 160 ---QGRYTTDDGYIFNASDIIEDTGAYIVPHGDHVIYIPKNELSAELAAAEFLSGRNL 217
DB 1121 AIQPAITTKD---NAQAIATKANERKTA-----IAQTQDITAEETAAANA-----D 1164

QY 218 LNSRTYRQNSDTSRTNWPVSNSPCTTNTSNNSNTNSQASQNDISLLKQYKYL 277
DB 1165 VDNVAVTQANSIEAANSQNDVDQAKTGETSID-QVTPVANKATARNEITALNN--KL 1221

QY 278 PLSQRHVESDGLVFPDPAQITRTARGVAVPHGDHVIYIPYQMSSELEERARIIPLYRS 337
DB 1222 QEIQATPDATDEEKQAADAEANTENGKA-----NQAIASAATTNAQVDEAKANA 1269

QY 338 NHHVPDSRPEQSPQPTPEPSPQPPA-----PNLKIDSNSSIVSQLVKVGEG--- 386
DB 1270 EAAINAVTPKVVKQAADKDEIDQLOATQTNVINDQATNEKEKAIQOLATAVTDKNN 1329

QY 387 ---YVFEKGISRYVFAKLPSETVKNLES-----KLSQESVSHLTAKKENVAPRQDEY 440
DB 1330 ITAAATDDNGVDLT---AKDAGKNSIGSTQPATAVKSNKNEVDQAVTTQNAI----- 1378

QY 441 DKAYNLLTEAHKALFXNKGNSDFOALDKLLER-INDESTN-----KEKLVDDLLAFLAP 494
DB 1379 DNTTGATTEE-----KNAAKDLVLKAKKAYQDILNAQTNDVTQIKDQAVADIQGITAD 1433

QY 495 ITHPERLGPNSQIEYTEDEVIAQLADKYTTSDGYIFDEHDIISDEGDVYTPHMGSH 554
DB 1434 TTIKD-VAKDELATKANEOKALIAQTADATT-----BEKQANQVDAQLT--OCNQ 1483

QY 555 WIGKDSLSKKEKVAQAAYTKEGILPPSPDADVKANPTGDSAAAIYNRVKGKRIPLVRL 614
DB 1484 IENASQIDVDNTAKONAI---QAIDPIQASTVDKTNARAELLTEMQNKI-----TEI 1532

QY 615 PYMVEHTVEKGNLIIP-----HKDHYHNKFAWF--DDHTYKAPNGYTTLEDLFATIKY 668
DB 1533 LNNNETTNEEK-GNDIGPVRAAYEEGLNINNAATTTGDVTTAKDTAVQKQVQQLHA----- 1586

QY 669 VEHPDERPHSNDGWGNASEHVLGKDHSDPNKKNKFADEE-PVEETP--AEPEVPOVETE 725

Db 522 GSNALRTEKGSVES-----NSMLQESNGEIIKCEAKENRQP----- 561
 Qy 315 IPIYSOMSELEERLARIILPLRYRS---NHWVPDRPEQPSPQTPPEPSPG----- 360
 Db 562 ---AKSFTEEAARIIQSVMRGVDVERWEPFKKLKEIA---TVREQMGVKKRIEALFA 615
 Qy 361 --PQAPNLKIDNSSILVQVRKVGEGYFEEKIGISRYVFAKDLPSFTVKNLESKLS-- 416
 Db 616 STDQHIEKEEIVNGELVNMILLKLD---AVEGLHPSIREFRKALATE-LSSIQDKLDSL 671
 Qy 417 KQBSVSHTLTAKKENAVAPROEFYDKAVNL---LTRAHKAFLXNKGKNSDFQALDKLLE 472
 Db 672 KNSCASAKEAVQEVELKSP-SDSPVNLBHSQLTENKVV----- 712
 Qy 473 RLNDSTNKEKLVLDLLAFLAPIITHPERLKGPNQIEYTEDVEVRIQADKYTSDGYIF 532
 Db 713 ---SDTNLEKVL---LSPEEH-----PMSVLNRTDEK---QAESAAATEEGY-- 751
 Qy 533 DEHDIISDEGDAYVTPHMGHSHWIGKDSLSKKEKVAQAAYT-----KEGILPPSPDA 585
 Db 752 ---PSADGNGMTVNVNENKAMVVESELEPEINELPQVBETETNSIRDPENASEVSEA 791
 Qy 586 DVKANPTGDSAAIYRNVKGEKRI-----PLVRLPYMVEHT-----VEV--- 624
 Db 792 ---PSADGNGMTVNVNENKAMVVESELEPEINELPQVBETETNSIRDPENASEVSEA 846
 Qy 625 -----KNGNLIIPHKD---HYHNKIFAWFDHT-----YKAPNGYTLDELFA--- 663
 Db 847 ETNSENENRKGEDIVLHSEKNVSELSPLVGVIDEETQPLSQDSSSYTREGNMTAMP 906
 Qy 664 -TIKYVVEHPDRPHSDNGWGNASHVLGKKHSDPNKNFKADEPVE-----ETPAE 716
 Db 907 KTASQEEETEVDHSPNKSIGIGQOTSEPODEKEQS--PETEVIVKQPLETEVILNEQAP 964
 Qy 717 PVPQVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLNNLTQLIMNNSIMAEAE 776
 Db 965 PEITEPGISKETKLMENORFETMETIVKAGREQ--LEVISKUITSRVKSLKLSHKK 1022
 Qy 777 KLLALLKGNPSVSVS 791
 Db 1023 KTQIRRRASKPWSVS 1037
 RESULT 16
 G89957
 hypothetical protein [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: G89957
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: G89957
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-891 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701526; PIDN:BAB42820.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA1552
 Query Match 4.0%; Score 165; DB 2; Length 891;
 Best Local Similarity 19.0%; Pred. No. 0.27;
 Matches 177; Conservative 140; Mismatches 323; Indels 292; Gaps 48;
 Qy 7 YQATVKNRNVSDKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTHGDHY 66
 Db 36 HQAQAENNTSDKISENQNNNATTQOPKDTNQTQ--PATQPV--ITAKNYPAA----- 86

Qy 67 HYNGKVPYDAIISSELLMKDPNYKLDKEDIVNEVKGYYVI---KVDGKYVYV--LKDA 121
 Db 87 -----DESLKDAI--KDPALENKEHDIGPREQVNFOLLKNNETQYHYFFSIKDP 135
 Qy 122 HADNVRTKEI-----NROK-----QEHSHQREGTPTNDG 152
 Db 136 DVYVYTKKAEVELDINTASTWKXFEVYENQKLPVRLVSYSVPVEDHAYIR---FPVSDG 192
 Qy 153 AVALARSGQRYTTDDGY-----IFNASDIIE-DTGDAIVPHGDHYHYPK 197
 Db 193 TQEL-KIVSQTQIDDEETNYDTKLFAKPLYNDSLVKSDTNDAAVVT----- 240
 Qy 198 NELSASLAAABAPLSGRGNLSNRTYRRQNSDNTSRNWP-----SVSNPG----- 245
 Db 241 NDQSSSDAS-----NQNTNT-SNQNTSTNNANNOPOATTMNSQAQKSSANA 289
 Qy 246 -----TTNTNTSNSTNSQASQSDIDSLLKQLYKLPQRSQVESDGLVFD----- 292
 Db 290 DQASSQPAHETNSNGNTNDKTNESNQSD-----VNOQYPPADESLQDAIKN 336
 Qy 293 PAQITSTARGAVPHGDHYHIFIPYSQMSLEERARIILPLRYSNHWWPDSRPEQSPQ 352
 Db 337 PAIDKEHT-----ADNWRPIDQMKDKGER-----QFYHYA----- 369
 Qy 353 PTPPSS-----PQGPAPNLK-----IDNSLSLVQLYR-KVGEGVVPEEKGIS 395
 Db 370 STVEPATVIFTGTGVEIHELKSTAKTWKKFVEYEGDKLPVELVSYSDSKDYAIRFPVS 429
 Qy 396 RYVFAKDLPSSTVKNLESKLSQBSVSHLTAKKENAVAPRQDEFYD-KAYN---LLTBAH 451
 Db 430 N--GTRDV--KNVSIEGIEGNIHEDYDTLMVFAQPIINNPDYVDEETYNLQKLLAPVH 485
 Qy 452 KALFXNKGNSDFQALDKLERLNDE-STNKEKLVDDLFLA-----PIHPERLKGPN 505
 Db 486 KA---KTLERQVVELEKLEKLPKRYKAEYKKLDQTRVELADQVKSATVEFENVPTN 541
 Qy 506 SQIYTEDEVRIAQLADKYTTS--DGYIFDEHDIISD--EGDAYVT--PHMGHSHWIGKDS 560
 Db 542 DQLTDLQEAHFVFESESESVMDGFV--EHPFYATILNGQYVVMKTKDDSYN--KDL 597
 Qy 561 LSDKEKVAQAAYTEKG---ILPPSPDADVKANPTGDSAAIYRNVKGEKRIPLVRLDY 616
 Db 598 IVEGHRVTVSKDPKNSRTLIFFYPD-----KAVYNAL---VKVWVANIGY 642
 Qy 617 MVEHTVEVKNGLIIPHKDHVHNI-----KFAWFD--DHTYKAQNGYTLDEL 661
 Db 643 EQQYHVRILNQDINTKDDTSQNTSEPLNVQTQGEKVADTTVAENSTATNPKDASDK 702
 Qy 662 FATIKYVVEHPDRPHSN--DWGNASEHVLGKKHSDPNKNFKADEPVEETPABPE 718
 Db 703 ADVIE-----PESDVVKDADNNDKQVHDVHLSMDSDNNHFD----- 741
 Qy 719 VPQVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLNNLTQLIMNNSIMAEAE 763
 Db 742 -----KYDLKEMDTQIAKTDNRV-----DNSVGMSSNVDTDKDSNKNKDKVIOLA 787
 Qy 764 QIMDNNSIMAEAEKLLALLKG-SNPSSVSKEK 794
 Db 788 HIADKNHHTGKAALDWVKQVNNNTDKVTDKK 819
 RESULT 17
 A44361
 amiloride-sensitive sodium channel Apx protein - African clawed frog
 N:Alternate names: apical plasma membrane protein
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 30-Apr-1993 #sequence_revision 19-Oct-1995 #text_change 22-Jun-1999
 C:Accession: A44361; S25517
 R:Staub, O.; Verrey, F.; Kleyman, T.R.; Benos, D.J.; Rossier, B.C.; Kraehenbuhl, J.P.
 J. Cell Biol. 119, 1497-1506, 1992
 A:Title: Primary structure of an apical protein from Xenopus laevis that participates in
 A:Reference number: A44361; MUID:93107151; PMID:1134959
 A:Accession: A44361

A:Molecule type: mRNA
A:Residues: 1-1420 <STA>
A:Cross-references: EMBL:Z14997; NID:G64551; PID:CAA78718.1; PID:g64552
A:Experimental source: A6 cells
A>Note: sequence extracted from NCBI backbone (NCBIP:121141)
C:Complex: This protein is part of a large molecular complex.
C:Function:
A:Description: may be the amiloride-sensitive component of the amiloride-sensitive sodium channel
C:Superfamily: amiloride-sensitive sodium channel Apx protein
C:Keywords: glycoprotein; membrane-associated protein; sodium transport
F:119,462,481,503,660,984,1038,1211,1273/Binding site: carbohydrate (Asn) (covalent)

Query Match 3.9%; Score 161; DB 1; Length 1420;
Best Local Similarity 19.88; Pred. No. 0.88;
Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

QY	42	EGINAEQIVIKITDQGVYVTSQGHV-----HYNGKVPY-----DAIISEE	82
Db	26	ERISPRVSMITLVDSAYSSFGSSGYVEYQNSFQHDGCHYNDQQLSYMDSEYVRAIYNS	85
QY	83	LMKDPNYKLKDEDIVNEVRGGYVIVKDGK-----YVYLKDAAHADNVRTK---	129
Db	86	LLDKDGVY-----NDIVSE-HGSSKVALSGRSSSILCSDNITTSVHRTSFAKLDNYVTNLD	140
QY	130	-----EIN-ROKQHSQHREGTTPND--GAVALARSQGR-----YTTDDGYIFNA	173
Db	141	EKNYIGDPIINMKHKONRPNHKAYLQNSPTGINSLOEKENQLYNPSNFMFKIDNYFGRS	200
QY	174	SDIIETDGAIVPHGDHYH---YIPKNELSAELAAAEFLSGRNLNSR---TYR	225
Db	201	LDVLQADGD---IMTQDSYTONALYFPQNPQDYR---NTQYFANRMSKEQFKNVDVQ	253
QY	226	RQNSDNTSRNMVPSNPGTTNTNNSNTNSQASQSDIDSLIKOLYKLPLOSRHVE	285
Db	254	KGNEENTERDG--PYLTKDQGF-----VOGYASDVRTSPKNI-----RRSLK	294
QY	286	SDGLVPDPAQIITSRTARG---VAVPHGDHYH-----IPYSQMSGELEERII---	331
Db	295	KSA-----SGKIVAHDSQGSQWIMKPGKDTSPNSEGTITMDYDNREQWDIRKSLSTRA	350
QY	332	--PLRYRNHHVPDSRPEQSPQPTPEPSPGPQAPNLKIDSNSLSVLVKVGEGVVF	389
Db	351	SOSLYYESNEDV-----SGPELKAMNSKNEVDQTLN-----	381
QY	390	EKGISRYVFAKDLPSYTNKLESKLSKQBSVSHTLTAKKNVAPRQDFYDKA-----Y	444
Db	382	-----FKQDATVKSIPLLSQQLQOEKCKSHPL-----SDLNCEKITKASTPMLY	425
QY	445	NLLTEAHKALP-----XNKGNSDFQALDKLLERLNDESTNKEKLVDDLAFAPITHPE	499
Db	426	HLAGGRHSAFIAPVHNTNPAAQEFKLESKTLERMNIS-----VLQSEPRPDNH	476
QY	500	RLGKNSQIEYTEDEVRIQAQADKYTTSQGYIFDEHDIISDEGDAYVTPHMGSHWTGKD	559
Db	477	KLPKNKS-----LTQLADLHDSVEG-----GNSGNLSSABE	508
QY	560	SLS-----DKEKVAQAAYTKEGKI-----LPPSPDADVKANP-----	591
Db	509	SLANDYIEKLKVAQKVLRETSEKFKDLQMSLPCREKLNPPKPTIDHFRSYSSSSANEE	568
QY	592	-----TGDSAAAIY-----NRVKEKRIPLVRLPYMVEHTVEVKNGNLIIPH-K	635
Db	569	SAYLQTKNSADSSYKKDDTDEKVAVTRIGRKR-----TKEQKLCYSEPEKLD	617
QY	636	H--YHNIFKAFDDHTYKAPNGYTTLEDLPAITKYVEHPDPRHSDNGWGNASHEVLGKK	693
Db	618	HLGTQKSNPAWKEETFANRREMSODISANRIKYLE-SKERTNSS---SNLSKTELKQI	673
QY	694	DH-----SDPNKNKFKADEPVEETPA-----EPEVPQVETEKVAQLK----	732
Db	674	QHNALVQWMERKTQRPNSN---PQVQWERTSLGLPNYNWNSLYSETSSSDASQYLRR	730
QY	733	-----EAEVLLAKVT-----DSSLKANATETLAGLNRLTLQIMDNNSNIMAEAKLLAL	781

QY	459	GRNSDFQALDKLERLN---	DESTNKEKLVDDL---	LAFIAPT---	THPERLGKPN	505
Db	2065	IREQDRKQVETVLROVEI	PPSDDLNNQSNTHPDKV	KVAVIELARLWPLYFNQV	VEVVEKRPD	2124
QY	506	SQIE---	YTEDVRIA---	QIAKYTTSQGYIFDEHDI	ISDEGDAYVTPHMGSHWI	556
Db	2125	ESVSTIFAI	SEHGIRLIVTHPHDLEN	PLKIQDFPFETIADV	SLSEANDILSVHVRH---	2180
QY	557	GKOSLSDKVKVAAQYK	TEK---	GILPSPDADVKAN---	PTGDSAA	597
Db	2181	-----	EDEENAYSARV	IKTNQAPQIKYTLOR	CLSGGVPRKRVLARTGEI	PPVCN--- 2232
QY	598	AIYNRVKGEKRIP	VLRLPYVEHTVEYKNGN	LIIP-----	HKOHYHNKIPAWEDDHTYKA	652
Db	2233	WLYGKIEN---	-----	RGFLLAQVYD	TDCGNVPPIRHETSE	DRDVRFP-- 2283
QY	653	PNGVTLEDL	FATIKYVVEHPDERPHS	NDGWCNASEHVLGK	KDHSEOP--	KNKFADEEPEVE 711
Db	2284	SERYTMID-	FAT-KYFRKPKDKKQ	ETWAMEDISQIV---	RFSKPI	SOSLIADLIGNEE 2337
QY	712	ETPAEPEVPQVETE	KVAQAQKEAE-----	VILLAKVTDS	SSLKANATETLAGR	NNLTL 763
Db	2338	SKYAVEIFHAI	MKEWGDLEPKK	SESMTDVVFKV	LLICHROPTLR--	DEVVCOLIKQTTTS 2394
QY	764	QIMDNNSTMAE	AKLALLKKGNSP	SSVS		791
Db	2395	NISOKPNSAL	RAWRLITITAIYF	SSSLT		2422

RESULT 19

T18444
 hypothetical protein C0385C - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18444
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18444
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1650 <LAW>
 A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
 C:Genetics:
 A:Map position: 3
 A:Introns: 1597/3, 1625/3
 A:Note: C0385C

Query Match	3.8%;	Score 158;	DB 2;	Length 1650;	
Best Local Similarity	17.7%;	Pred. No. 1.6;			
Matches	146;	Conservative 127;	Mismatches 311;	Indels 242;	Gaps 34;
Qy	64	DHYHYNGKVYDAIISEELMKDPNYKLKDEDIVNEVKGWYIVKDGKYYWYLKDAHA	123		
Db	11	DNVHYTKQV-----IRYDHNNIKANSHLKQSEY-----TTTH	47		
Qy	124	DNVR-----TKEEINROKQHSQHREGGTPRNDGAVA-----LARSQGYT	164		
Db	48	DEKQWITSNNNTNNNIQTHA-----NVKDDKSVENKTTTLINKDINEITWISKQDNYT	102		
Qy	165	TDDGYIFNASDIIEDTDGAYIVPHGDHYIYIPKELASASELAABAFITSGRGLNSSTY	224		
Db	103	KYNCLNFNKSVDVWRD-----YVYRSKKEDSSNNMGNNGNNMGNMGNMNNM	150		
Qy	225	RRQSDNTSRTNWVSVSNPGTTNTNTSNNSTNSQASQSDNDISLLAQLYKLPLSQRHV	284		
Db	151	NGNNNN-----GNNMNGNNMNGNNMNDNNNG-----NNI	185		
Qy	285	ESDGLVEDPAQITTSRT--ARGVAVPHGDHYHPIYQSQSELEERIARIPLEYRSGNHWPV	342		
Db	186	MNGNNMNGNNNINRNNIMNGNNMGN-----NINRNNIMVR-----NNITNNHHI	234		
Qy	343	DSRPEQSPQPTPEPSGQPAPNL-----KIDNSLSVQ-----	378		

Db	235	HI SNKTS LIENNKKKEENI PPSANLYKKEMNVKGSSDTFDLLYKRKIDDDNNLSKKKRE	294
Qy	379	--LVRYKGEVYVPEKGISR-----YVFAKOLPSTVTNLSKLSKO	418
Db	295	NULFNKDGHEFVNKENYQNNIIDDODDDDDNNHDNVVVYDKVKENENKKNKYSKE	354
Qy	419	ESVSHTLT--AKKENVAPRD--OEFYDKAYNLLTEAHKALFXNKGSRNSDFQALDKLLER	473
Db	355	DGLNVLVELNKNLNVNDNIINKSPKNNILYIKTSDSLNENYNERKIYKEINK----	410
Qy	474	LNDESTNKEKUV-----DDLAFIAPITHPERLKPNSQIETEDDEVRIAQLADKYTSD	528
Db	411	--EYSNKNYVYHFKNND-----SSIKKNNSSCELDQKKKTY---KYTIE	454
Qy	529	GYIPEHDIIDSGDAYVTPHMGSHWIGKDSLSDEK-----VAAQYATKEGILPP	581
Db	455	QKRYNFD--RDNNNAYI-----KDDTHKKEGYLLNMIYQSEY--KTYGSNNK	500
Qy	582	SPDADVKANPTGD-----SAAATYRVKGEKIPLVLPYVMEHTVEYQNG--	627
Db	501	MDMEIYNGHTNDFNINENLNNKIYDDYEGYDPEKKKKRLD-----DHIYTOQEKYNNI	556
Qy	628	NLIIPHKDHYHNIKPAWFDDHTYKAPNGYTTLEDLFATIXYVYHPDPERPHSNDGWNASE	687
Db	557	NDIL--KDHLLN-----DKETEKKEKEIIEE--EKKNKIEIBBEKKNKIE-----IEE	601
Qy	688	HVLGHKHSDEGNKPKADEEVEETPAEPVQVETEK-----VEAQLK	732
Db	602	EKKNKIEIEEKKKIEIEEKKKIEEKKKIDDEKKNYANDKIISHDYNVNCIK	661
Qy	733	BAEVLAKVTSSLSKANATETL--AGLRNNLTLQIMDNNSIMAPAEK	777

RESULT 20

A:Accession: S73361
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-910 <H1M>
 A:Cross-references: EMBL:AE000004; GB:U00089; NID:gl673671; PIDN:AAB95683.1; PID:gl6736
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 C:Genetic code: SGC3
 C:Superfamily: Mycoplasma heat shock protein C09 homolog C09_orf910; dnaJ amino-termin
 C:7-71/Domain: dnaJ amino-terminal homology <DNJ>

Query Match	3.8%	Score 157;	DB 1;	Length 910;
Best Local Similarity	19.0%;	Pred. No. 0.77;		
Matches 161;	Conservative 127;	Mismatches 305;	Indels 254;	Gaps 44
QY	66	YHYNGKVPYDAITSEEL-----LWKDP-----NYKLKDEDIYNEVKGGYVIKVD--GKYI 114		
DB	34	YHPDRNKAPDRAQIPAEINEANDVLNSNPKRANYDYKYGHDGVDN--EGGFQADVPDSFF 92		
QY	115	VYLKDAAHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARCGRYTTDDGYIFNMS 174		
DB	93	BEIEKSGAFDNL--SESNTKKXKTKTKKKGW-----FWGKSQKEESTSDTTEY--A 140		
QY	175	DIIEDTGDAIVPHGDHYHYIPKNELSAELAAAEFLSGRNLNSRYRQNSDNTSR 234		
DB	141	DV--DAGLEDYPPQSDYDDIPDVPDARIEVDQS--AYADDIPDVGMDWEQNAEVAUNIA 197		

QY 791 SKEKN 796
| : |
Db 1439 EAEKN 1444

RESULT 24
 G84598
 Probable bZIP transcription factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84598
 F:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.J.; Town, C.D.; Pujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84598
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-519 <STO>
 A:Cross-references: GB:AE002093; NID:G4582442; PIDN:AAD24827.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g21230
 A:Map position: 2

Query Match	3.7%;	Score 156;	DB 2;	Length 519;	
Best Local Similarity	19.4%;	Pred. No. 0.39;			
Matches 125;	Conservative	84;	Mismatches 176;	Indels 260;	Gaps 28;
QY	246	TTTNTSNTSNTNSQSQSDNDISLLKQLYKLPQSQRHVSDGLVDFDPAQITSRARGVA	305		
Db	7	TTDTNMQRVNSSGTSSTSS-----IPKHNHL-----NPALIRS-----	41		
QY	306	VPHGDHYHF-----IPYSOM-SELEBRIARII---PURYRSNHWVDSRPE	347		
Db	42	-----HHHRHPPTGAPPPPIPIPSYQIPATLQPRHGRSMSPQSPFFSDFSLPLPINS	96		
QY	348	QPSQPQTPPE-----SPGQPAPNLKIDSNSSLVSQLVRKVGEGYVFBEKGISR-----	396		
Db	97	AFPSVSVVEEKTGAGFSPSLPPSPPTMCHSSSS-----RNAGDGENLPKRKHRRNSD	150		
QY	397	-----YVFAKDLPSTVTKNLSKLSQBSVSHTLTAKKENVAPRQDEFYDKAYNLLTEA	450		
Db	151	VTFGSSMMNSQKSPPLSLRSISGEDTSDWSNLVKKE---PR-EGFY-----	196		
QY	451	HKALFANKGRNSDFQALDKLLERLNDESTNKELVDDLLAFAPITHPERLCKPNSQIBY	510		
Db	197	-----KGR-----KP-EVEA	205		
QY	511	TEDEVRIAGLADKYTTSDGYI FDEHDIISDEGDVYVTPHMGSHWIGKDSLSDKEKVAQA	570		
Db	206	AMDDVFTAYM-----LONIDVLSFG-----GEDGKNGNENVEEMESSRGS	247		
QY	571	AYTKEGILPPSPDADVKANPTGDSAAALYNRVKGEKRIPLVRLPYMVBHTVEVKNGLI	630		
Db	248	GTKTKTNG--GSSSDSEGSDSSAGNVKVALSSSSSGVKR-----RAGGDI	289		
QY	631	IPHKDHYHNI-----KFAWFDHDTYKAPNGYLTLEDLFATIKYVVEHPDBRPHS---	678		
Db	290	APTGRHYRSVWMSDFMGKLNFGDESSLLKP-----PSSSAKV	327		
QY	679	---NDGWGNASEHVL---GKKDHSEDPNKNFKADEBPVEETPAEP-----	718		
Db	328	SPTNSGEGNSSAYSVEFGNSEPTAEMKKIADEKLAELVMDPKRVKRIILANRVSAARS	387		
QY	719	-----VPQVETEKVEAQLKEAEVLLAKVT-----DSSLK--ANATETLAGL	757		
Db	388	KERKTRYMAELE-HKVQTLQTEATTLTSLQALTHLORDSMGLTQNQNSSELAFRQJAMEQQOAL	446		
QY	758	RNNLTLOIMDNNSINAAEAKLLAL-----KGSNPSPSVSKEKIN	796		
Db	447	RDALUSEKUNE-----EYORLKLVTICEPNRRQSGSSSESMSLN	485		

RESULT 25

T18416
 Hypothetical protein C0105w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18416
 R:Lawson, D.; Bowman, S.; Barrrell, B.
 submitted to the EMBL Data Library, July 1997
 A:Reference number: Z18934
 A:Accession: T18416
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1338 <LAW>
 A:Cross-references: EMBL:Z97348; NID:el323671; PID:el323672; PIDN:CAB10568.1
 C:Genetics:
 A:Map position: 3
 A:Note: C0105w

 Query Match 3.7% Score 154.5; DB 2; Length 1338;
 Best Local Similarity 16.6%; Pred. No. 1.8;
 Matches 153; Conservative 122; Mismatches 287; Indels 357; Gaps

Query Match	3.7%	Score 154.5	DB 2	Length 1338
Best Local Similarity	16.6%	Pred. No. 1.8		
Matches 153	Conservative 122	Mismatches 287	Indels 357	Gaps 35
QY	11	TVKENNVSYIDGKQATQKTENL	-----TPDEVSKREGINABQIVIKITDQGYVTSH	62
Db	277	TQNKDKNIEY	:-DQKCTSKENIEDNVSFVNDPSDPNQKNLNN-----NITDNNIIPSN	329
QY	63	-----GDHYHYNGKVPYDAIISBELLMKDPNYKLKDEIDVNEVKGGVYKVDGKYVYVLK	118	
Db	330	VQIEKQSTLSKKNKNEKDSYINNSLTNDQNLKREDIKENDKAEGITKYD-----MLNLIK	386	
QY	119	DAAHADNVRTKEI	-----NRQOEHSQHREGGTPRNDGAVALARSGRYTTDDGYI	170
Db	387	-----NNISIKINCHSPNENKKNKDNHQ	-----CEDNSI	418
QY	171	FNAASDIIEDTGDYIIVPHGDHYHVIIPANELSASELAAAEAFLSGRGNLSNSRTRYRONS	230	
Db	419	-----	-----NICN-----NKN	425
QY	231	NTSRTWVSVSNPGTTN	-----TWTSN-----NSNTNSQASQSDI-----DSLKLQOLY	275
Db	426	NNIQTNNI	-----NDNTVNEKINNTSKMDLNNLTQNNNDSEKNDVWIEQQLVNEIDLKKN	481
QY	276	KLPLSRHVESDGLV	-----FDPAQITSRARGVAVPHGDHYHIFIPYSQMSLEERI	327
Db	482	KQTKKKKNINEPYVKHKLSPNSDSLTS	-----YSNIHALQETL	523
QY	328	AR	-----IIPLYRSN-----HWVPD-----SRPEQSPQPTPEPSPGQPAPNLK	368
Db	524	TRKPHYNTYFLNNPEKYRDNKMPVHLRLPNDCLKKIDQDSDEFEEEDDLSVDONKE	583	
QY	369	IDSNS	-----SLVSQLVKVGEGYVFEKGTISRVAKDLPSETV-----KN	410
Db	584	QKNQOLEVNLPNKYPNSNDVYKFPFKDINKFPPIYCDMFNHLTHPEALRLHELYMKUKKN	643	
QY	411	LESKLSKQESVSHTLTAKENVAPDQE	-----FYDKAYNLLTEAHKALPKNKG	459
Db	644	IDSNNMTDLGNQNSHKVYIINTEDGEYICIRPYDPVYHYHEKSCYKICDLGNSLWIDS	703	
QY	460	RNSDPQA	-----LDKLLERLN	475
Db	704	RYAEIOTRQYRAPEVILKSGFNETADINWSFACWVFLVTGDFLWPKQGDYRDKNEHLS	763	
QY	476	-----DESTNK	-----EKLVDOL	488
Db	764	FLIEVLGNI PKHMIDAGVNSHKYFNKNYRLKNIRNIKKYGLYKILKYKNLPEKEISPL	823	
QY	489	LAFLAPIHTHBERLGKPNSEQIYTE	-----DEVRIAQLADKYTTSDDGYIFDEHDIISDGG	542
Db	824	CSFLLPMLSDVPQTPPSA	-----YTMQLHPWLNWVSLEEGDDVYNDESYISINDRNKNNS	880
QY	543	DAYTTPHMGHSHWTGKDSLSKEKVA	-----QAYTKEXG-----	577
Db	881	NSNNFIYDGHNSRRKNKS	-----SNKKKIDVNYKILGNNGNNAIYNDNYKNKYKNKNKNKNKNFNDD	939

us-09-765-271-56.rpr

Fri Oct 1 15:35:55 2004

Qy	578	ILPSPDADVKANFTGDSABAIIYNRVKCEKEIPLVRLPYVVEHTVEKNGNLIIPHKHCHY	637
Db	940	VVEFSPDQYMHANYNNDIVHAV-----LYEFPYNNNVI SYN-----NKCHK	982
Qy	638	HNIKFAWF-----DDHTYKAPNGYTLDFATIKYYVEHPDERPHSNDGNGNASEHVLGK-	692
Db	983	NNFDINYLOHRNDNNSNKONISLTNDYTFNSDYIANMMDHDTYRKQIKNIPAHQISKL	1042
Qy	693	KDHSDDPNKFKADEEPVE	711
Db	1043	KD-----GKNFKAYNESIQ	1056

Search completed: October 1, 2004, 07:17:16
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 03:36:05 ; Search time 19 Seconds
(without alignments)
2181.465 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTKENRVSY.....KLLALLKGNPSSVSKEKIN 796

Scoring table: BLOSUM62

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 90 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	4.1	1233	1 YF16 YEAST	P43597 saccharomyc
2	161	3.9	1420	1 AXP_XENLA	Q01613 xenopus lae
3	158	3.8	1118	1 UBPF8 HUMAN	P40818 homo sapien
4	157	3.8	910	1 DNJM MYCPN	P75354 mycoplasma
5	157	3.8	1658	1 YM67 YEAST	Q03661 saccharomyc
6	157	3.8	2195	1 SC16 YEAST	P48415 saccharomyc
7	156.5	3.8	1185	1 MAPX DROME	P23226 drosophila
8	154.5	3.7	3924	1 ANK2 HUMAN	Q01484 homo sapien
9	154	3.7	1453	1 Y373 BOVIN	Q9TU23 bos taurus
10	153.5	3.7	1539	1 Y373 HUMAN	O15078 homo sapien
11	153.5	3.7	2867	1 RBP2 PLAVB	Q00799 plasmodium
12	150.5	3.6	1176	1 YOH8 YEAST	Q08236 saccharomyc
13	149.5	3.6	1175	1 PTNL RAT	Q62728 rattus norv
14	149	3.6	2869	1 REP1 PLAVB	Q00798 plasmodium
15	145.5	3.5	1085	1 CUT7 SCHPO	P24339 schizosacch
16	145.5	3.5	1240	1 YNJ1 YEAST	P53935 saccharomyc
17	144	3.5	1337	1 JIP3 MOUSE	Q9EN9 m c-jun-ami
18	143.5	3.4	900	1 SVA MYCGE	P47534 mycoplasma
19	142.5	3.4	875	1 Z1P1 YEAST	P11111 saccharomyc
20	142.5	3.4	1066	1 STS5 SCHPO	O74454 schizosacch
21	142	3.4	850	1 D7 DICDI	P54682 dictyosteli
22	142	3.4	1744	1 TANA XENLA	Q01550 xenopus lae
23	141.5	3.4	490	1 MOT3 YEAST	P54785 saccharomyc
24	141.5	3.4	1928	1 MYSL YEAST	P08964 saccharomyc
25	141.5	3.4	2314	1 PTPZ HUMAN	Q00402 saccharomyc
26	141.5	3.4	2748	1 NUM1 YEAST	P23471 homo sapien
27	141	3.4	914	1 IP42 YEAST	Q00401 saccharomyc
28	141	3.4	1332	1 SPT7 YEAST	P39336 saccharomyc
29	140.5	3.4	516	1 P54 ENTFC	P13692 enterococcu
30	139.5	3.3	1609	1 LMGI HUMAN	P11047 homo sapien
31	139	3.3	1957	1 SPOF SCHPO	Q10411 schizosacch
32	138.5	3.3	1703	1 SNF2 YEAST	P22082 saccharomyc
33	138	3.3	1639	1 MSPI PLAFW	P04933 plasmodium

34	137.5	3.3	1790	1 USO1 YEAST	P25386 saccharomyc
35	137.5	3.3	2175	1 HMCU DROME	P10180 drosophila
36	137.5	3.3	2459	1 MABP RAT	P15205 rattus norv
37	137	3.3	1176	1 PTNL MOUSE	Q62136 mus musculu
38	136.5	3.3	679	1 YIS3 YEAST	P40563 saccharomyc
39	136.5	3.3	1183	1 CNA STAAU	Q53654 staphylococ
40	136	3.3	1630	1 MSPI PLAFK	P04932 plasmodium
41	136	3.3	1836	1 BUD3 YEAST	P25558 saccharomyc
42	135.5	3.3	1053	1 SLPM BACBR	P06546 bacillus br
43	135.5	3.3	1376	1 MYHA BOVIN	Q27991 bos taurus
44	135.5	3.3	3358	1 PGCW MOUSE	Q62059 mus musculu
45	135	3.2	1532	1 IGA NEIGO	P09790 neisseria g
46	135	3.2	3664	1 MINT HUMAN	Q96458 homo sapien
47	134.5	3.2	960	1 YMX6 YEAST	Q04279 saccharomyc
48	134.5	3.2	1124	1 TCF8 HUMAN	P37275 homo sapien
49	134	3.2	1163	1 RTN4 RAT	Q9J111 rattus norv
50	134	3.2	1188	1 OSH1 YEAST	P35945 saccharomyc
51	134	3.2	6632	1 UNB9 CAEEL	P01761 caenorhabdi
52	133.5	3.2	1607	1 LMGI MOUSE	P02468 mus musculu
53	133.5	3.2	1664	1 INT1 CANAL	P53705 candida alb
54	133	3.2	1189	1 YOH6 YEAST	P47035 saccharomyc
55	132.5	3.2	1875	1 MLF1 YEAST	Q02455 saccharomyc
56	132	3.2	1164	1 TSC1 HUMAN	Q92574 homo sapien
57	132	3.2	2431	1 POLN SFV	P08411 senliki for
58	131.5	3.2	764	1 MPS1 YEAST	P54199 saccharomyc
59	131.5	3.2	1046	1 YU14 YEAST	P47029 saccharomyc
60	131.5	3.2	1379	1 YFP9 SCHPO	O14066 schizosacch
61	131	3.1	1014	1 HEX2 YEAST	Q00816 saccharomyc
62	131	3.1	1093	1 TWF1 HUMAN	P82094 homo sapien
63	131	3.1	1358	1 SIR4 YEAST	P11978 saccharomyc
64	131	3.1	1612	1 DNM1 PARLI	Q27746 paracentrot
65	130.5	3.1	704	1 MSN2 YEAST	P33748 saccharomyc
66	130.5	3.1	1547	1 TOP2 BOMMO	O16140 bombyx mori
67	130.5	3.1	1849	1 IGA4 HAEIN	P45386 haemophilus
68	130.5	3.1	2805	1 MAPA HUMAN	P78559 homo sapien
69	130	3.1	969	1 SABC STRSL	Q55242 streptococ
70	130	3.1	1283	1 OSH2 YEAST	Q12451 saccharomyc
71	130	3.1	1829	1 RM1 MOUSE	O70472 mus musculu
72	130	3.1	6669	1 NEBU MOUSE	P20929 homo sapien
73	129.5	3.1	1001	1 RFRG MOUSE	Q910X5 mus musculu
74	129	3.1	646	1 SG1 BOVIN	P23389 bos taurus
75	129	3.1	1026	1 STAÜ DROME	P25159 drosophila
76	129	3.1	1031	1 YDG9 SCHPO	Q10496 schizosacch
77	129	3.1	1570	1 P3K1 DICDI	P54673 dictyosteli
78	129	3.1	2278	1 FABI YEAST	P34756 saccharomyc
79	129	3.1	3038	1 TRIO HUMAN	O75962 homo sapien
80	128.5	3.1	1137	1 MSB1 YEAST	P21339 saccharomyc
81	128	3.1	742	1 CD44 HUMAN	P16070 h cd44 anti
82	128	3.1	817	1 YG4A YEAST	P46949 saccharomyc
83	128	3.1	1039	1 MSL1 DROME	P50535 drosophila
84	128	3.1	1616	1 P200 MYCGE	Q49429 mycoplasma
85	127.5	3.1	782	1 Y044 UREPA	Q9PRA1 ureaplasma
86	127.5	3.1	954	1 BIR1 YEAST	P47134 saccharomyc
87	127.5	3.1	1334	1 JIP3 HUMAN	Q9UPT6 homo sapien
88	127.5	3.1	1466	1 SPA2 YEAST	P23201 saccharomyc
89	127.5	3.1	1597	1 GTF1 STRDO	P11001 streptococ
90	127.5	3.1	1608	1 HLYA SERMA	P15320 serratia ma

ALIGNMENTS

RESULT 1	YF16 YEAST	STANDARD;	PRT; 1233 AA.
ID	YF16 YEAST	AC	P43597;
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Hypothetical 137.7 kDa protein in UGS1-FAB1 intergenic region.		
GN	YFR016C.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Maitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae."
 RL Nat. Genet. 10:261-268 (1995).
 CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D50617; BAA09255.1; -
 DR PIR; S56271; S56271.
 DR GermOnline; 140170; -
 DR SGD; S0001912; YFR016C.
 KW Hypothetical protein.
 SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

 Query Match 4.1%; Score 171; DB 1; Length 1233;
 Best Local Similarity 19.8%; Pred. No. 0.15;
 Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;

 QY 12 VKENRVSIVDGRQATQKNTLTPDEVSKREGNABQIVI-----KITD-QGVVTS 61
 DB 151 IKETSTNNVAEG-----TENVP--IKSTGIEVGNPSITRRKKNKKKTTNRRGRNS 202
 QY 62 HGDHYHYNGKVPYDAII--SELLMKDPNKLKDEI-----VNEVGGYVTKVD 110
 DB 203 NPADTTDLSCQSTLSDLSILVIEEYQLQEDGS---KNEDIKVINVDPEVNVKEMDIRTNE 259
 QY 111 GKVVYVLKAAHADNV-----RTKEENRQKQHSQHREGG--TPRNDGVALARSO--- 160
 DB 260 SSKTFDIDVKNKDNDVETSKSENNEINEKAHTLPRNEILNVNENAGNAAPFKHQLP 319
 QY 161 -GRYTTDDGYFNASDIIEBTGDYIVPHGDHYHYPNLSASELAAPLSGRNLS 219
 DB 320 HGLEAGDE---NGQASTKDVESRLTKNGFNFKENESKHLKAGEKQOTE---SDRDGIS 372
 QY 220 NSRTYRQNSDNTSRTNWVPSVNP-----GTINTNTSNNSTNSQASOSNDI----- 267
 DB 373 PSVLAKNQKTEIGKEDHVFQKDEKCKRKLSDVNNHNNMNFNAAGSDSIIPPETE 432
 QY 268 -----DILLKQLYKLPLSQRV-----ESDGLVFPDPAQ---ITSR 299
 DB 433 REYVDETMGTPTKISDNEKNLQHTNDISVEVEKEEEEEENSTFSKVKKENVTGE 492
 QY 300 -----TARGVAVPHGDHYHYPYQMSSELEERTARILPLYSRN----- 338
 DB 493 QEAVRNNEVSGTEESTSKGEIMGGD-----EKQSEAGEK-SSIIETEGSANSKIS 544
 QY 339 --HWVPSRPPQSPQPTPEPSPGPQAPNLKIDNSLSVQVKGVEGVVFEKGISR 396
 DB 545 KDNVLLEDAEAPTOENKPTVEVGEIDIPDPRDD-----VEIVAEVKNILPELVAK 599
 QY 397 -----YVPAKDLPSVETKNLSKQSVSHHTITAKENVAPR-----DOEFYDK 442
 DB 600 EDQEGEQVKLDEPVKMKDDKIAMEGASISSEDMKKKQEGTAELSNEKAKKEVDETARES 659
 QY 443 AYNLLTEAHKALFXNK-----GRNSDFQALDKLLELNDESINKEKLVDDLLAFLAPI 495
 DB 660 AEGVEVEKSKTPSPKPVKVRCTSGRPDLQINERDPEVLKEDVRVPDPEIATTIEN 719

QY 496 THPERLKGKNSQIEYTEDEVRIAQLADKYTTSGDGYFDEHD-----IISDEGDVYTPHM 550
 DB 720 SEEDPKSORVQISTEQAETTKQMDGVGTSSTTS---FKEEKPKRFEITQEGDKITGKT 776
 QY 551 GHSWIGKDSLSDEKKAQAAYTKKGILPPS-----PDADVKNPTGDSAAIYNR 602
 DB 777 NHEHGEATEAASSENSK-ASDVGTAEKYTFPSSSVKVKDTEEDAEE-----NSEKTEFIK 830
 QY 603 VKGKRIPLVPLPYVVEHTVEVKNGLII-----PKDHYHNIKFAWFDHTY 650
 DB 831 VRAE--LENLDAPKAEVTAELNKENEDVEVDTEEDAEEVENSEKTEFIKVK---AELGNL 885
 QY 651 KAPN--GYTLE-----DLFATIKYVVE---HPDERPHNDGNGNASEHVLGKD-- 694
 DB 886 DAPKAEVTAELNKENEDVEVAATSKEDTETKSEPAETPIED---GTCTEAEVSKKDAE 942
 QY 695 --HSEDPNKNFKADEPVEETAEPEVPOVE-----TEKVE-AQLKEAEVLLAKVTDSSL 746
 DB 943 AVTKEDENMENSKIABALKDVTGDQDIDINISDEFQRTVELPELEKQDIKNGKDEKEL 1002
 QY 747 KANATETLAGL-----RNNLTILQIMDNNSIMAEKLLALLKGSNPSVSKKIN 796
 DB 1003 EVEETEKETSLPDLVVEENIT---BEKNIKQEEEB-VSQLDNFNETESISKEAPN 1053

 RESULT 2
 APX_XENLA STANDARD; PRT; 1420 AA.
 ID APX_XENLA
 AC Q01613;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Apical protein (APX).
 GN APX.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 MEDLINE=93107151; PubMed=1334959;
 RA Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,
 RA Kraehenbuhl J.-P.;
 RT "Primary structure of an apical protein from Xenopus laevis that
 RT participates in amiloride-sensitive sodium channel activity."
 RL J. Cell Biol. 119:1497-1506 (1992).
 CC -!- FUNCTION: Is part of a multimeric complex which is involved in
 CC amiloride-sensitive sodium channel activity.
 CC -!- SUBCELLULAR LOCATION: Membrane-associated.
 CC -!- TISSUE SPECIFICITY: Kidney, proximal intestine, oocytes, and
 CC to a lesser extent in the distal intestine, stomach and eye.
 CC -!- SIMILARITY: SOME, TO HUMAN APXL.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z14997; CAA78718.1; -
 DR PIR; A44361; A44361.
 KW Membrane; Sodium transport; Transport.
 FT DOMAIN 111 114 POLY-SER.
 FT DOMAIN 559 564 POLY-SER.
 FT DOMAIN 1048 1051 POLY-THR.
 SQ SEQUENCE 1420 AA; 159467 MW; 0DD8B5C11413FFBC CRC64;

 Query Match 3.9%; Score 161; DB 1; Length 1420;
 Best Local Similarity 19.8%; Pred. No. 0.62;

Matches	180;	Conservative	111;	Mismatches	318;	Indels	302;	Gaps	44;
QY	42	EGINAEQIVIKITDQGYVTS	HGDHY	----	HYNGKVPY	----	DAIIEE	82	
Db	26	ERISVRSMTTLVDSAYSS	FGSSVPEYQNSFQHDGCHYND	EQLSYNDSEYVRAI	YNPS	85			
QY	83	LLMKDPNYKLEDEIVNEKGYIK	VDGK	----	YVYVYLKDAHADNV	RTK	----	129	
Db	86	LLDKDGVY	----	NDIVSE-HGSSKVALSGRSS	SSSLCSDDNTTSVHRTSPAKLDNTV	TNLDLS	140		
QY	130	-----	EBIN-RQKEHSQHREGGT	PRND--GAVALARSQGR	-----	YTTDDGYIFNA	173		
Db	141	EKNIVGDPINMKHKQNR	HNKAYGLQRNSPTGINSLOEKENQLYNPSNFWEIKDNY	FGRS	200				
QY	174	SDIITDGDYIVPHGDHYH	----	YIPKNEUSASELAAAEFLSGRGNLSNR	----	TYR	225		
Db	201	LDVLAQDGD	----	INTQDSYTONALYFPON	PDQYR----	NTQYFGANRMSKEQPKVNDVQ	253		
QY	226	RQNSDNTSRTNWVPSVSNPGT	TNTNTSNNSNTNSQASQSDNIDSLLLKQLYKLPLSQRHVE	285					
Db	254	KSNEENTERDG	----	PYLTKDQGF	-----	VOGYASDVRTSFNKI	----	294	
QY	286	SDGLVFPDAQITSTARG	----	VAVPHGDHYHF	-----	IPYSQMSSELEARI	II	----	331
Db	295	KSA	----	SGKIVAHDSQGSQWIMKPGDKDTFSN	EGTITDMDYDNRREQWDIKRSL	STRA	350		
QY	332	--	PLURYSNHWVPDSRPEQSPQTPPE	SPQPPAPNLKIDSNLSIVSLQVRKVGEGYV	F	389			
Db	351	SQSIIYESNEDV	-----	SGPFLKAMNSKNEVD	DTLS	----	381		
QY	390	EEKGISRYFAKOLPSETVKNLES	KLKQBSVSGHTITAKKENVAPRQDFYDKA	----	Y	444			
Db	382	-----	FQKDATVKSI	PLLSQQLQEKCKSHPL	-----	SDLNCEKIKTASITPMLY	425		
QY	445	NLLTEAHKALF	-----	XNKGENSDFOALDKLLRLNDESN	TKELVDLLAFIAPITHPE	499			
Db	426	HLAGGRHSAFIA	PVHNTNPAQOEK	KLKLESTLERMNIS	-----	VLQLSEPRPDNH	476		
QY	500	RLGKPNISQIEYTEDVRIAQLADKY	TTSDGVI	FDEHDIISDEGDAYVTPHMGHSHWIKD	559				
Db	477	KLPKNS	-----	LTQLADLHDSVEG	-----	GNSGNLSABE	508		
QY	560	SLS	-----	DKEKVAQAQYTKKEGI	----	LPSPDADVKAMP	----	591	
Db	509	SLMNDYIELKVAQK	VLRETSPFKRDLQMSLPCRFKLP	PPKRPPTIDHFRSYSSSSANEE	568				
QY	592	-----	TGDSAAAIY	-----	NVKEGKRIPLVRLPYWVHTVEVNGNLI	IPHK-D	635		
Db	569	SAYLQTNASDSSYK	KDDTEKVAVTIGKRR	-----	TKEQKKLCYSEPEKLD	617			
QY	636	H--	YHNKIFAWFDDHTYKAPNGY	TLEDLFATIKYVVEHPDERPHSDNGMGNASEHVLGKK	693				
Db	618	HLGIQKSNFAWKBEPT	FANRREMSDSDISANRIKYLE-SKERTNS	----	SNLSKTELQI	673			
QY	694	DH	-----	SEDPNKNFRADEP	VPETPA-----	EPEVPQVETEKVBAQIK	----	732	
Db	674	QHVALVOYMERKTNQRPNSN	----	PQVMERTISGLPLNPNYENWISY	SETSSDASQYLRR	730			
QY	733	----	EAEVLLAKVT	-----	DSSLKANATETIAGLRNLLTIQIMDNNSIMAEKLLAL	781			
Db	731	RSAGASSSYDAITV	WNDRFGKTSPLGRSAE	KTAGVQRK	----	TFSDQRTLDGSGEH	----	783	
QY	782	LKGSNFPSSVSK	792						
Db	784	LEGSSPSLSQK	794						

RESULT 3

UBP8 HUMAN

ID

UBP8_HUMAN

STANDARD;

PRT; 1118 AA.

AC

F40818;

DT

01-FEB-1995 (Rel. 31, Created)

DT

01-FEB-1995 (Rel. 31, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 8) (Ubiquitin-specific processing protease 8)
DE (Deubiquitinating enzyme 8).
GN USP8 OR KIAA0055.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II."
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
CC -! CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2O) =
CC ubiquitin + a thiol.
CC -! SIMILARITY: Belongs to peptidase family C19.
CC -! SIMILARITY: Contains 1 rhodanese domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).

EMBL; D29956; BAA06225.1; --
DR MEROPS; C19.011; --
DR Genew; HGNC:12631; USP8.
DR MIR; 603158;
DR GO; GO:0004497; F:cysteine-type endopeptidase activity; TAS.
DR GO; GO:0004843; F:ubiquitin-specific protease activity; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR InterPro; IPR001394; Peptidase_C19.
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR Pfam; PF00443; UCH; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANESE_3; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_1.
DR PROSITE; PS50235; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT DOMAIN 195 313 RHODANESE.
FT ACT_SITE 786 786 BY SIMILARITY.
FT ACT_SITE 1059 1059 BY SIMILARITY.
FT ACT_SITE 1067 1067 BY SIMILARITY.
SQ SEQUENCE 1118 AA; 127523 MW; 8B884B7A842F9A9A CRC64;

Query Match 3.8%; Score 158; DB 1; Length 1118;
Best Local Similarity 19.0%; Pred.No. 0.64;
Matches 173; Conservative 143; Mismatches 336; Indels 258; Gaps 45;

QY 6 LYQAARTVKNRVSYDYGKAQTOKNTPTDVSKREGINAEIQIVIKTDGQYVTSHGDH 65
Db ||| :||: |:
QY 11 LYLSSSLKDLNK-----KTE-VKPEISTKSYPHSALKIFKTAEERCLDR--DE 56
Db ||| :||: |:
QY 66 YHYINGKVYPDALISELLMKDPNKXEDDEVNEKGYYIVKDGKYVIYVKDA--- 121
Db ||| :||: |:
QY 57 ERAVLVMKYTVYN--LTKKPPDFK-QOODFYHSILPGNIKKAVEEAERLSISKLR 113
Db ||| :||: |:
QY 122 HDADNVRTK-EEINRQK-EHSQHREGGTPRDNGAVALL----- 156
Db ||| :||: |:
QY 114 EEAEVRKKLEEKROEEAQRLQQEQETGREDDGTLAGSLENULDSDKDKTKQSKEKNE 173
Db ||| :||: |:
QY 157 ---ARSQRYSITDDGYTF----NASDI-----EDTGDAVIPHGHDHYIPKNELSA 203
Db ||| :||: |:

Db 174 KCETKEGAIITAKELYTMTDKNISLIIMDARMQDYQDSCL-----HSLSVPEEAI SPG 229
 Qy 204 ELAA-AEAFSLGRGNLSRTYRRQNSDNTSRNNVPSVN--PQTTNTNTSNNNTNSQ 260
 Db 230 VTASIEAHL--DDSDTKWKRGNVEYVLLDFSSAKDLQIGTTLRSKL-----278
 Qy 261 ASQNDIISLLKQLPLSQRHVESDG-----LVDPDAQITSRARGVAVPHGDHYHFI 315
 Db 279 -----DALFWESKTVLRNEPLVLEGGYENWLLCYQYTTNAK--VTPP-----320
 Qy 316 PYSQMSLEERATARIPLRYRNHNVDPDRPQSPQPTPEPSQPQAPNLKIDNSSL 375
 Db 321 PRQ-----NEEVSISLDFTPS--LEESIPKPAQ-----TPPASIEVDENIEL 364
 Qy 376 VSQLVRKV-----EGYVFEKGISRYVFAKPLPSTVKNL-----ES 413
 Db 365 ISQNERMGLNISTPVEPVAASKSDVSPII-----QPVSINKVPOIDRTKPAVKLPPE 420
 Qy 414 KLSQESVSHLTAKENAPDOEYDKAYN--LLTEAHK-----ALEXNKGRNSD 463
 Db 421 HRIKSESTNHEQSQSGKVPIDRSTKPVVFGPTLMTDEEKARIIHAETALLMEKNKQB- 479
 Qy 464 FOALDKLLRLNDESTNKEKLVDDLLAFLAPITHPERLQKPNISQI---BYTEDEVRIAQ 520
 Db 480 -----KELRERQEE--QKEKLKE-----EQEQKAKKKQEAENEITEKQKAKEE 524
 Qy 521 ADKYTTSQGYIFDEHDIISDEGDAYVTPHMGHSHWIGKSLDKKVAQAAYTKEGILP 580
 Db 525 MEKKESEQAKEDKE--TSAKRKEIT-----GVKROSKSHSETSDAKSVEDRGKRC 575
 Qy 581 PSPDADVKA-----NPTGDSAAIYNRVKGKRIPLVPLPYMVVHTVEVKNL-----629
 Db 576 PTEIOKSGDGVHSTVSITGDSGKPFKIQGPESGIILRTGTFREDTDTTERKQAQREP 635
 Qy 630 -----IIPKDHVHNIKFAWDDHTYKAPNGYT--LEDLFATIKY-----VEH 671
 Db 636 LTRASEEMGRIVP-----GLPSGWAFLDPITGTFRYHSPNTVH 677
 Qy 672 -----PDRPHSNDGWNASHVHLKGDHSDPNKNFKADEEP-----VBETPAEPE 718
 Db 678 MYPEMAPSAPPSTPTTHKAKTQIPABRD--REPSKLKESYSSPDITQAIQEEKKRKT 735
 Qy 719 V-PQVETEKVEAQLKAEVILLAKVDTSSLK-----ANATETLAGLNR-----NLTLQ 764
 Db 736 VTPVNRNKPCTYPKAEI--SRLSAQIRNLNPLVFGSGFALTLGLNLGNTCYMNSILQ 793
 Qy 765 IMDNNSIMAE 774
 Db 794 CLCNAPHAD 803

RESULT 4
 DNJM MYCPN STANDARD; PRT; 910 AA.
 AC P7535a;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DnaJ-like protein MG200 homolog.
 GN MPN119 OR MP035.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae."
 RL Nucleic Acids Res. 24:4420-4449 (1996).
 CC -!- SIMILARITY: Contains 1 J domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000004; AAB95683.1; --
 DR PIR; S73361; S73361.
 DR HSP; P25685; 1HDJ.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008971; HSP40 DnaJ_pap.
 DR InterPro; IPR003095; Hsp_DnaJ.
 DR Pfam; PF00226; DnaJ_1.
 DR PRINTS; PR00625; DnaJPROTEIN.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS00776; DnaJ_2; 1.
 KW Hypothetical protein; Chapterone; Complete proteome.
 FT DOMAIN 4 73 J-DOMAIN.
 SQ SEQUENCE 910 AA; 100190 MW; 125D0E37D2D221A7 CRC64;
 Query Match 3.8%; Score 157; DB 1; Length 910;
 Best Local Similarity 19.0%; Pred. No. 0.55;
 Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;
 Qy 66 YHYNGKVPYDALISEL-----LMKDP-----NYLKDSDIVNEVKGVIKVD--GKYV 114
 Db 34 YHPDRNKAPDAQIPAEINEANDVLSNPKKRYNYKXGHDGVN--EGGFAPQADVDSFF 92
 Qy 115 VYLKDAHADNVRTKEEINRQKESQHREGGTPRNDGAVALARSGRYTDDGYTFNAS 174
 Db 93 ERIEKSGAFDL--SESNTKKKTKTKKGW-----FWGSKQESTSDTTEY--A 140
 Qy 175 DIIEDTGDYIVPHGDHYHVPKVELSASELAFAFLSGRGNLSRTYRRQNSDNTSR 234
 Db 141 DV--DAGLEDYPPQSDYDDIPDVARIEEVDQS--AYADDIPDVGDMQWQAEVANS 197
 Qy 235 TNWVPSVNGTNTNTSNNNTNSQASNDISLL-----271
 Db 198 SEIIPDV-----DAGLADEFNTSSAPQASDWEAMIGNPEYGYFDAAGWNKGFDEA 251
 Qy 272 -KQLYKLPLSQRHVESDGLVDPQAITSRTARGVAVPHGDHYHFIPIYSQMSLEERIAI 330
 Db 252 GQWVLEETEPSSVNDITDSDAVTAAT-----VEETDQ-----288
 Qy 331 IPLRYRNHVPDSRPE-----QPSQPTP-----EPSPGP-----OPAPNLKI 369
 Db 289 -----DSWTANSAPEPVDVETPVELQPEPEPIITLSSEPVEAPASVVIETPIEE 341
 Qy 370 DSNSSLVSQLVRKVGEGYVFEKGISRYVFAKPLPSETVKNLSKLSQBSVSHLTAKK 429
 Db 342 TTSAVEMQASVK-----ADVSDEADATNEPTEQUTISEPEQ--ETDAAALEINH-TADL 394
 Qy 430 ENVAPRDQEFYKAYNLLTE--AHKALFXNKGNSDFQALDKLLERLNDESTNKEKLVDD 487
 Db 395 E---PAEV---SATNDLEQDVVEKVFSEPESTVDTAATDPVVEQATETSTNGKFPF-N 446
 Qy 488 LLAPL-----APITHPERLQKPNISQI-----EYTEDEVRIAQLA-----521
 Db 447 FSSFVLSDQNPNPOTPTTHHEEADAAPETVDETSGESTAPEVTIETAEVLETAENNP 506
 Qy 522 -----DKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLS 562
 Db 507 ATFVEEVLQPTKTTVVDKLEDEPTVAKTVSDSENSVAPEPEFVAGPEQTFE-W--KPAIS 563
 Qy 563 DKEKV-----AAQAYTKEGILPSPDA-----DYKANETGDSAAA-----IY 600
 Db 564 ETEBIPLTAVEPASETQTLIAEDVTSFVTPTATAPAPSINAVPTAPVAETFEAAVDLCK 623
 Qy 601 NRKVGKRIPLV-RLPYMVVHTVEVKNLIIIPKDHVHNIKFAWDDHTYKAPNGYTL 659

```

Db 624 EAAKIEAQLPVTPTVEQIDGT-----DPSLL-----TQWDEYLEKTRK--LFH 665
QY 660 DLFAATIKYVVEHPDERPHSDGNGWASEHVLGKKHSE--DNKNPKADEEPEVEETPAP 717
Db 666 KLFLP-----EQLP-----FIYKTDQFEIVDEN-----LDEHVNLIYTE- 700
QY 718 EYPQVETKEVAQLKEAEVLLAKVDSSLKANATETLAGLRNLTLOIMDNNSIMAAEAK 777
Db 701 HVPQICP--LNEQLKEIR-YTRKLVDPQTVTTT-----SITLEVQLSHKSQTEA-- 748
QY 778 LLALLKG 784
Db 749 -IAIFKG 754

RESULT 5
YM67_YEAST
ID YM67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUA1-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin M., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome
RT XII."
RL Nature 387:90-93 (1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z49809; CAA89934.1; -.
CC PIR; S55101; S55101.
CC GerMOnline; 142894; -.
CC SGD; S0004832; ESC1.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0006348; P:chromatin silencing at telomere; IMP.
CC Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.8%; Score 157; DB 1; Length 1658;
Best Local Similarity 19.7%; Pred. No. 1.2;
Matches 176; Conservative 119; Mismatches 338; Indels 262; Gaps 41;

QY 21 IDGKAQTKTENITPD--EYKRE---GNAEQLVITKIDQGVTS-----HGDIHYH 68
Db 193 LEGKQSLIKFDLENDYELSEKESKSDGSSPSIMI--LSDEYAEAGALQDVSNDYAE 251
QY 69 YNGKVPYDAIISBELMKD-----PNY-----KLKDE-DIVNEVK----- 102
Db 252 EEEQVERKNIGQEQANVENATQISSDSSSGYQSEGVENELEDDIDVESDAKDSOGA 311
QY 103 GGVIVKDGKYVYLKDAADHADVTRKEINRQKQHSQHREGGTPRNDGAVALARSQGR 162

```

```

Db 312 EGTEHSVDFSKYM---QPRDTNTKIPVIEKYESDEHKVHQ-----R 349
QY 163 YTTDDGYIFNASDII-----EDTDGAYIVPHGDHYHVIKPNELASALAAAFLSG 214
Db 350 YSEDCAFDFGSVNI SVDDDESEDEESQAESYANAEVYHH-NEHLLDDKEL--IEDIES- 405
QY 215 RGNLSNSTYRONSNDTSRTNWFVSNGPGTTNTNTNNSNTNSQA-----S 262
Db 406 ---SDSESQAESQESQEDDFEYKMKNEKSTSETENTSESQDQGFADKAYTKNKVEQ 461
QY 263 QSNIDID-----SLLKQLY---KLPLSRHVESDGLVFDPA-----QITSRT 300
Db 462 QENDEEPEKDDIIRSLDKNPHGNKSEYSENVLENE---TDPALVRENOINDVEGYD 518
QY 301 ARGVAVPHGDHYH-----FIPYSQ-----MSELEERIIARIIPLY--RSNHW 341
Db 519 VTGKSVESDLHEHSPDNLVLAARAMLQFQOSRNSNCPCQKEEQVSE---SVLHGNSGN 574
QY 342 PDSRPEQSPQTPPEPSPQAPNPKIDSN--SSLSQLVKVGEGVVFEEKGISRYVF 399
Db 575 LSGRSLDESEEQIPLKQFTGENNNLNKTDGDLSSSVIEVEKVE----- 620
QY 400 AKDLPSETVKNL-----ESKLSKQESVSHTL-----TAKKE 430
Db 621 -KXLDGSTKELVPLSTDTTINNSLGNEDSIYSLDDADALISEMLTDVPLMEIKTTPKY 679
QY 431 NVAPRDOEFYDKAY--NLLTEAHKALFXNKGNSDFQALDKLLRLNDESTNKEKLVDDL 488
Db 680 EVWISSEVYSTSYEDNTVAMPQVEYTSPPMNDPFNS-----LNDDYEKKHDLAKST 732
QY 489 LAFLAP-ITHEP-----RLGKENSQIEYTEDVRI-----AQLADKYTSDGYIFDHD 536
Db 733 LAALAPFTKKDAEFVAGVTKSLTSTSGHTNIFHTSKETKQVSDLESTENVTF--ENE 791
QY 537 IISDE--GDAYVTPHMGHSHWIGKDSLSDKEKVAQAQYTKEGILPPSPDADVKANPGD 594
Db 792 NTGDNKNQSKNFPFVANSTDKSTEDNTDEKYFSAINT-----NVTGD 835
QY 595 SAAIYNRVKGEKRIPLVRLPYMVEHTV-----EVKNGNLIIPHKD--HYHNIK 642
Db 836 SSC------DIETASNVEENLYCEKDMNEAMSSGDECVKQNDGSKTQISF 884
QY 643 AWFDDHTYKAPNGVTLDELATIKYVVEHPDERPHSDGNGWASEHVLGKKD--HSEDPN 700
Db 885 STQSPDNFQSNNDNT---EFSSTKYKVRNSDLEDESLKXELTKARVVDKLDDEESEDY 941
QY 701 KNFKADEEPEVEETPAEPE-----VPQVETEKVEAQLKEAEVLLAKVTD 744
Db 942 EQDYADPEPGNDEGSNENIVKGTGKOTLGIVEPENKVKVHEETLFEANVSS 996

RESULT 6
SC16_YEAST STANDARD; PRT; 2195 AA.
AC P48415; Q02822;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Multidomain vesicle coat protein.
GN SEC16 OR YPL085W OR LPF1W
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017704; PubMed=7593161;
RA Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that
RT interacts with Sec23p."
RL J. Cell Biol. 131:311-324 (1995).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN-S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansonge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Borstein D., Bowman S., Brucker M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hallier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidic A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.,
RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI";
RL Nature 387:103-105 (1997).
CC -!- FUNCTION: Involved in the budding of transport vesicle from the
CC endoplasmic reticulum. The C-terminal interacts with SEC23 and
CC with the cytosolic domain of SED4. Could therefore be a
CC constituent of copII vesicle coat. N-terminal overexpression
CC causes a lethal secretion defect.
CC -!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
CC WHICH BUD FROM IT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL; U23819; AAC39088.1; -;
CC EMBL; U41849; AAB68254.1; -;
CC PIR; S61103; S61103;
CC GeneOnline; 144067; -;
CC SGD; S0006006; SEC16.
CC GO; GO:0030127; C:copII vesicle coat; IDA.
CC GO; GO:0005789; C:endoplasmic reticulum membrane; IDA.
CC GO; GO:0005198; F:structural molecule activity; IPI.
CC GO; GO:0005914; P:autophagy; IMP.
CC GO; GO:0016192; P:vesicle-mediated transport; IDA.
CC Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
CC DOMAIN 1997 2094
CC MUTAGEN 1059 1059
CC L-S: IN SEC16-4; TS ACCUMULATION OF ER
CC MEMBRANES
CC L-S: IN SEC16-3; TS ACCUMULATION OF ER
CC MEMBRANES.
CC L-S: IN SEC16-2; TS ACCUMULATION OF ER
CC MEMBRANES.
CC W-S: IN SEC16-1; TS ACCUMULATION OF ER
CC MEMBRANES.
CC MISSING (IN REF. 1).
CC I -> F (IN REF. 1).
CC CONFLICT 522 522
CC CONFLICT 560 560
CC SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;
SQ

Query Match
Best Local Similarity 19.7%; Pred. No. 1.8;
Matches 124; Conservative 109; Mismatches 241; Indels 156; Gaps 31;
QY 188 HGDHYHYPKNELSASELAEEAFLSGRGNLSNRTYRQNSDNTSYNWPVSNGTT 247
Db 29 HSEPLEPESTINSS-----FNDDSVNRTESDIASKSD-VPPVVS-----S 68
QY 248 NTNTS-----NNSNTNSQASQNDI-----DSLKLKLYKLPL 279
Db 69 STNISSPANETOLEPTDQELHKLKNDQDIDITADSGNDIPNDSIVHDSVITQT-KPAM 127
QY 280 SQRHVESDGLVFDAQITSR-----TARGVAVPHGDHYHFFIPYSQMSE----- 322

Db 128 SQEYEET-----AAHLSRNPSPSLDVAGELHNNNEHTQKIAVSAAVEEDSFNEEGENHD 181
QY 323 --LEERIIARILPLRVSRNHWPDSPRQPSQPTPEPSPGPOAPNLKIDNSSLVSLV 380
Db 182 SIIISLNDATPSQY--NHLFSPD-----GNLSPSLSGDTPTHNVLPTGKDN----- 228
QY 381 RKGVGYYFEEKGISRYVFAKD-LPSETVKNLESKLSQESVSHLTITAKENVAPRDQBF 439
Db 229 -EINDEVCNDKEIS--LNANNVLPDELSEKEDERLKELETHVS--TEEEKQDIA--DQET 281
QY 440 YDKAYNLLTEAHKALFXNKNRSD--FOALDKLLERLNDESNKLEKLVDDLLAFIPIH 497
Db 282 AENLFTSSTPESENKIRNSGDTSMLFQ-----DDESDQKVPWEEDVKDF----H 328
QY 498 PRLGKPNISOIYETDEVK-----IAQLADKYTTSDGYIPE-----HDIISDEGDAYVTP 548
Db 329 NENTNNTQESAPNTDORDKGYEGNALKKSECTAADERSYSEETSEDIIFGHDKQVVE- 387
QY 549 HMGSHHWIGKDSLSDKEKVAQAAYTKKGIILPPSPDADV--KANPTGSAIAIYNRVKGE 606
Db 388 --GQNDFTGKNLENSQKLMGSGNHK---LPLSAEADIIIEPKQIQDQADLFTQSSGD 441
QY 607 KRIPLVRLPYMVEHTVEVKNGLIIPHKDHYNI-----KFAW-----FDDHTYKAPN 654
Db 442 -----LGEVLPWESTDKNADVTSQSEKHEDLFAASGNDEKLPMEVSDSGEVSSGKTEN 494
QY 655 G-YTLEDLPATIKY-YVEHPDERPHSNDGWNASHEVLGKDKHSDSEPNKFNKADEPVEE 712
Db 495 SWQSTTEKIAEQKFSELENDLDDDDSLASSE-----EEDTVPTNTDNTNLTSPVEE 550
QY 713 TPAEPPEVPOVETEKEVAQKAEVLLAKVT 742
Db 551 KKAARYKPIIEE---EAGMRQGVHFTNTT 577

RESULT 7
MAPX DROME STANDARD; PRT; 1185 AA.
AC P23226; Q9V9S1;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 205 kDa microtubule-associated protein.
GN MAP205 OR CG1483.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91115949; PubMed=1703540;
RA Irminger-Finger I., Laymon R.A., Goldstein L.S.B.;
RT "Analysis of the primary sequence and microtubule-binding region of
RL the Drosophila 205K MAP";
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM B3).
RX STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.N., Cawley S., Dahlke C., Davenport L.B., Davies P.,

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kratt C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Mazny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiner B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 CC
 CC -!- FUNCTION: May play an important role in the regulation of
 CC microtubule assembly and interaction.
 CC -!- SUBCELLULAR LOCATION: Associated with cytoplasmic microtubules and
 CC with the mitotic spindle.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=J5;
 CC IsoId=P232226-1; Sequence=Displayed;
 CC Name=C2;
 CC IsoId=P232226-2; Sequence=VSP_004319, VSP_004320, VSP_004321;
 CC Name=B3;
 CC IsoId=P232226-3; Sequence=VSP_004319;
 CC -!- MISCELLANEOUS: Phosphorylation of various serine residues may play
 CC a regulatory role. The basic domain contains numerous sequences
 CC that match known consensus sequences of several different protein
 CC kinases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54061; CAA37996.1; --
 CC EMBL; AE003780; AAF57214.1; --
 CC PIR; A36685; A36685.
 CC FlyBase; FBgn0002645; Map205.
 CC GO; GO:0005875; C:microtubule associated complex; IDA.
 CC Microtubule; Alternative splicing; Phosphorylation.
 CC DOMAIN 1 784 ASP/GLU-RICH (ACIDIC).
 CC 785 1124 ARG/LYS-RICH (BASIC).
 CC DOMAIN 1125 1185 ASP/GLU-RICH (ACIDIC).
 CC BINDING 745 977 MICROTUBULES (POTENTIAL).
 CC VARSPLIC 557 578 Missing (in isoform B3 and isoform C2).
 CC /FTId=VSP_004319.
 CC VARSPLIC 650 703 Missing (in isoform C2).
 CC /FTId=VSP_004320.
 CC VARSPLIC 704 704 D -> N (in isoform C2).
 CC /FTId=VSP_004321.
 CC SEQUENCE 1185 AA; 126669 MW; 47B432E2CEE03F70 CRC64;
 CC
 CC Query Match 3.8%; Score 156.5; DB 1; Length 1185;
 CC Best Local Similarity 20.1%; Pred.No. 0.84;
 CC Matches 163; Conservative 111; Mismatches 304; Indels 231; Gaps 36;
 CC


```

Db 3194 TVQTGDIPPLS--GVKQISCPDSSSEPAVQVQL--DFSTLTSRVY-----SD----- 3235
QY 292 DPAQITSTRTAGVAVPHGDHVFHIFPYSQMSLEERARIIPIRYSNHWVDPSPREPQ-S 350
Db 3236 -----RQDDSPDSS-----PEBQKSVIE-----IPTAPMENVPFTESKSKIPVR 3274
QY 351 PPTPEPSPGPOPAPNLKIDNSSLSVLSQVLRKVGEGY---VFEEKGISRYVFAKDLPSRT 407
Db 3275 TMTPTSTPAP-----PSAEVSS-----VSEDFLSSVDDEENKADAPKSKLPVKV 3319
QY 408 -VKNLESLSKQESVSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGRRN-SDFQ 465
Db 3320 PLORVEQQLSDLD-----TSVQKTVAPQGDMAIAPD-----NRKSSESAS 3362
QY 466 ALD-----KLLERLNDESTNKEKLVDDLLAFAPL-----PVKSRSTTS 3418
Db 3363 SLDSKTKCPVTRSTETETESRERAELESEEGATRPKILTSRL-----PVKSRSTTS 3418
QY 496 -----THPERLGKN-----SQIBYTE-----DEVRIQALADKYTSDGVIFDEHDIISDE 541
Db 3419 SCRGTSPTKESKEHFFDLRYNSIEFFFEISDEA--SKLVRLTQSE-----REQEIVSDD 3472
QY 542 GDVYTPHMGSHWIGKSLDSKEXVAAQAQYKTEKGILPSPDADVKANPTGDSAAAIYN 601
Db 3473 -----EPPVQVETEKVQAQKBAEVLL-----AKVTDSS 745
QY 602 RVKGEKRIPLVRLPMVVEHTVEVKNGLIIIPKDHYNHNIKFAWDDHTYKAPNGVYLEDL 661
Db 3479 EVSVIENLP-----PVEHSHVP-----EDL 3499
QY 662 FAT-----IKYVVEH--PDRPHSNDGWNASEHVLGKDKHSEDPNKNFKADEPVEET 713
Db 3500 FDRPIMDESJETLIERIPDENG-----DHAEDP-----QDEQRIER 3539
QY 714 PA-----EPEVQVETEKVQAQKBAEVLL-----AKVTDSS 745
Db 3540 LAYIADHLGFSWTELARELDFTPEQIHQIRIENPNSLDQSQYLLKTVLWLRDGHATDTN 3599
QY 746 L-----KANA-----TETL-----AGRNLTLOIMDNNISMAEKLALL 782
Db 3600 LVECLTKINRMDIVHMLMETNTEPLQERISHSYAEIEQTITLDHSEGFSLQOE-BLCTAQH 3658
QY 783 KGSNPSSVSKE 793
Db 3659 KQKEEQAVSKE 3669

```

RESULT 9

```

Y373 BOVIN
ID Y373 BOVIN STANDARD; PRT; 1453 AA.
AC Q9TU23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RA Jovov B., Ripoll P.J., Benos D.J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF176816; AAF00990.1; -.
 KW Hypothetical protein; Coiled coil.
 FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF4139 CRC64;

Query Match 3.7%; Score 154; DB 1; Length 1453;
 Best Local Similarity 18.5%; Pred. No. 1.5;
 Matches 171; Conservative 124; Mismatches 329; Indels 298; Gaps 40;

```

QY 7 YQARTVKENRVSYIDGKQA-----TQKENTL---PD-----PD-----EVSKEGINAQOI 49
Db 486 WDQREVELEROLDVDFDROOSEILREAOKEFEATGSMPPDPLIPNQLEIALRK--IKENI 543
QY 50 VIKITDQGVYVSHGDHVVYNGKVPYDAIIIEELMKDPNPKYKDEDI-----VNEVKG 103
Db 544 RILLETQATCS-----LEEKREKESALRLAEENILSRDKVINEUR- 585
QY 104 GYVIKVDGKYVYVYLKDAHADNVRTKEINRQKQEHSHQREGGTPRNDGAVALARSGRY 163
Db 586 -----LRLPATAEQEKLLAEFSRKEVEPKSHHTLKLAHQ-----TIANMQARL 628
QY 164 TTDGDIYNASDIIEDTGD---AYIVPHGDHVIYI-PKNELSASELAAAEAFSLGRNLS 219
Db 629 NQKEEVKKYOHLEKAREEQREIVKKBHEELHTLHRKLELQA----- 671
QY 220 NSRTYRRONSNT---SRTNW---VPSVNSPTGTTNTNTSNNN-TNSQASNDIDSL-- 270
Db 672 -----DNSLSKFKETAWDLIKQSPVTPVFNKHFIRLAEMEQTVAEQDSSLSLVI 721
QY 271 LKQL-----YKLPUSQRH-----VESDGLVDFPAQITSR 299
Db 722 KLVQVQDLERKEITELKIKEFENMKLRQENHAEVKKIKAEVEDLRCLLVQSQESQ 781
QY 300 TARGVAVPHGDHVFHIFPYSQMSLEERARIIPIRYRS-----NHWVPDSRPEQSPQTP 355
Db 782 SLKSELOTQKANSRAPTTTNRNLVERLKSQALKEKQKALSALLLEAEAMTAABER 841
QY 356 EPSGPOPAPNLKIDS-----NSSLSVQLVRKVGEGYVEFEKISRYVFAKDLSETVKN 410
Db 842 IISWTSQKEANLVQIIVDRHTKELSQI-----EDL-NENILK 879
QY 411 LESKLSKQESVSHTLTAKKENVAPRQDFYD--KAYNLLTEAHKALFXNKGRRNSDFQALD 468
Db 880 LKEALKTSKNRENTLT---DNLDLNTNELQNKQKAYG-----KVLREKDAVDQENNELK 930
QY 469 KLLERLNDE-----STNKEKLVDDLLAFAPITHPERLKGPNKNSQIETEVDEVRIQALD 522
Db 931 RQIKRLTSGLGKPLIDNKQSLIEEL-----QKKIKLESQLEKRVDEAEKMPKE 981
QY 523 KYTTSDGYIFDE-----HOIISDEGDAYVTPHMGSHWIGK--SLSDREKVA 568
Db 982 KSAREEVIRWEEGKKMQTKTIEGRNKLKEGEVYI---LTQLTTLTKDLPAKADKELT 1038
QY 569 AQAYTEKGIPLPPSPADAVKANPTGDSAAAIYNRVKGEKRIPLVRLPMVVEHTVEVKN 628
Db 1039 LQRLKLTGTL-----TVQVMAA-RVLESEKSELE-----ELKRN 1072
QY 629 LIIPKDHVHNKIFAMFDDHTYKAPNGYITLEDLFATIKYVVE--HPDRPHSNDGWNAS 686
Db 1073 LDLENDISYMRSHQA-----LPRDSVIEDLHLQNKYLQEKHLALEKQLSKDAYSRPS 1124
QY 687 EHVLGKKDH-----SEDPNKNFKADEPVE-----ETPAEPEVPOVET----- 724
Db 1125 TSGIDSDHVVQREQLQRENKLSSNIELKFOLEQANKDLPRLKNQVRLKEMCEFLKK 1184
QY 725 EKVEAQKAEV-----LAKVTDSSILKAN-----ATETLAG 756
Db 1185 EKAEVERKLGRVGRSGRSGKTIPELEKTIGLMKVVEKQVRENEQLKASGILTSEKMAN 1244
QY 757 LRNNLTLOIMDNNISMAEAKL 778

```



```

Db 1245 IE-----MENEKLAELK 1259
RESULT 10
Y373 HUMAN
ID Y373 HUMAN STANDARD; PRT; 1539 AA.
AC O15078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB002371; BAA20828.1; -
KW Hypothetical protein; Coiled coil.
FT DOMAIN 18 1514
FT SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;
SQ
Query Match 3.7%; Score 153.5; DB 1; Length 1539;
Best Local Similarity 19.2%; Pred. No. 1.7;
Matches 200; Conservative 145; Mismatches 334; Indels 361; Gaps 53;
QY 3 ELGLYQARTVKNRVSIDGKATQKTEN-----LTPDEVSKREGINAEQIVIKITD 55
Db 306 EQALYYAR-LEGRNRAHL--RQTQSLRRQFSGALPLAQQKFSKTMQLQNDKLIKIM- 361
QY 56 QGVVTSHGHHYNGKVPYDAIS--EELL--MKDP-----NYKLKDEDI-VNEVKG 103
Db 362 QEMKNSQOHRNMENKTELEWELKGLGELISTLKTKGAQKVINHWKTEBELRLQELKL 421
QY 104 GYVIVKDGKYYVYLKDAHADNVRT---KEENRQKQHSQHREGTTPNDGAVALAR 159
Db 422 NRELVDKGEELNNII-SEVERTTSSLEEEIVQONKPFHEERQMAWDQRE---VDLERQ 477
QY 160 QGRYTTDDGYFNASDIED-TGDYIVPHGDHVVHPKNEI----- 200
Db 478 LDIFDRQNEILNAAQKFEATGS---IP--DPSLPLP-NOLTEALRKIKENIRILETR 531
QY 201 -----SASELAFAAEFLSGRGLNSRTYRRQNSDNTSRTNVPVS-----NP 244
Db 532 ATCKSLEELKESALRLAEQNILSRDKVINE--LRLRLPATAEREKLAELGRKEMEP 589
QY 245 GTTNT-NTSNNTNGQASQNSDIDSLKLYKPLPSORHVESDGLVDFDPAQITSTRARG 303
Db 590 KSHHTLKIAHQTIANMQA-RLNQKEEVLYKQYRILLEKAREEQ-----RE 632
QY 304 VAVPHGDHVFIPYSQMSLEERARIIPURYRSNHWVPDSRPQSPQPT----- 354
Db 633 IVKHEEDHLHLHLRLQADSSLN-----KFKQTAW--DLMQSPFPVPTNKHFIPLAE 695
QY 355 PEPSPGPAPNLKIDSNSSVLSQVLR-----KVGEGYVPE 390
Db 686 MEQTVAEQD-----DSLSSLLVLKLVKSQDLEROREITELKVKEFENIKLQLOQENHED 739
QY 391 EKGISRYVFAKDL-----PSETVKNLESKL----- 415
Db 740 VKKVAEV--EDLKYLLDQSKESQCLSELQAQKEANSRAPTTTMRNLVERKLSQLALK 797
QY 416 -SKQESVSHIL-----TAKKE--NVAPRDQBFYDK----- 442
Db 798 EKQKALSRALLEPRAEMTAAAEERIIISATSKAEHLNV-----QQIVDRHTRELKTQVED 853
QY 443 -AYNLLTEAHKALFXNKG-----NSDFQ--ALDKLL-----ERLND----- 477
Db 854 LNNELL-KLKEALKTSKNRENSLTNDLNDLNNELQKQKAYNKILREKEEIDQENDELKR 912
QY 478 -----STNKEKLVDDLLAFAPITHEPRLGKPNQSOLEYTEDEVRIQALADK 523
Db 913 QTKRLTSGLOKPLTDNKKSLIEEL-----QRKVKLENQLEGKVEEVDLKPMEK 963
QY 524 YTTSDGYIFDE-----HDIISDEGDAYVTPHMGHSHWIGKD--SLSDKEKVA 569
Db 964 NAKELIRWEEGKKWQAKIEGIRNKLKEGEVFTLTQQLNTL---KDLFAKADKEKLT 1020
QY 570 QAYTKEKGLPPSPDADKANPTGDSAAIYNRVKGEKRIPLVRLPYMVEHTVEVKGNL 629
Db 1021 QRKLKTTGM-----TVQVLGI-RALESEKELE-----ELKKNL 1054
QY 630 IIPKDHVHNIFAFDHDHTYKAPNGYLTDLFATIKYVVE--HPDERPHSNDGMCNASE 687
Db 1055 DLNDILYMRHQA-----LPDSDVVEDLHLQNYVLOEKHLEKQFSKDTYSKPSI 1106
QY 688 HVLGKKKH-----SEDPNKNFKADEEPVE-----ETPAEPVPOVET-----E 725
Db 1107 SGIESDDHCQRQELQKQENKLSSENIELKFQLEQANKDLPRLKNQVRDLKEMCEFLKKE 1166
QY 726 KVEAQLKEAEV-----LLAKVDTSSLKANATETLAGRNNTLTIQIM- 766
Db 1167 KAEVQRKLGHRGSGSGSKTIPELEKTIGLMKKWEKVQREN--EOLKKASGILTSEKWA 1224
QY 767 ----DNNSIMAEAEKLLALL 782
Db 1225 NIEQENEKUKAELEKUKAHL 1244
RESULT 11
ID RBP2_PLAVB STANDARD; PRT; 2867 AA.
AC Q00799; Q9NZM3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PvRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10838229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares
RT structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa
RT rhoptry protein family.";
RL Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -|- FUNCTION: involved in reticulocyte adhesion. Specifically binds to

```

human reticulocyte cells.
 --!- SUBCELLULAR LOCATION: Membrane-bound (Probable).

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF184623; AAF76525.1; --
 DR HSSP; P03069; IGCM.
 KW Malaria; Receptor; Signal; Transmembrane; Repeat.
 FT SIGNAL 1 21
 FT CHAIN 22 2867
 FT DOMAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.
 FT TRANSMEM 2806 2826 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2827 2867 POTENTIAL.
 FT DOMAIN 44 133 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 560 758 ASN-RICH.
 FT DOMAIN 1112 1285 LYS-RICH.
 FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.
 FT REPEAT 2758 2761 1.
 FT REPEAT 2762 2765 2.
 FT REPEAT 2766 2769 3.
 FT REPEAT 2770 2773 4.
 FT REPEAT 2774 2777 5.
 FT REPEAT 2778 2781 6.
 FT REPEAT 2782 2785 7.
 SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

 Query Match 3.7%; Score 153.5; DB 1; Length 2867;
 Best Local Similarity 18.5%; Pred No. 4.1;
 Matches 167; Conservative 143; Mismatches 348; Indels 247; Gaps 39;

 QY 12 VKENRV-SYIDGKQATQKTENLTPDEVSKR-----EGINAEQIVIKITDQGYVTS--- 61
 DB 663 ISKNDEIQKVI-----EKIEKLINDAPSGDKFTTEKTNLQNKVKIIDFHKEDLQLL 716
 QY 62 -----HGDIHYHN-----GKVPYDAI-----ISELLMKDPNY 90
 DB 717 LNSLSKPYEEHQLYNEASTIEKIKDLHQKTKEEYKLEKMKFSNFGQILDKLNTLNDL 776
 QY 91 KLKDEDIVNEKGGYVIVKVDGKYVYVYLKDAADANVRT-----KEENRQ 135
 DB 777 KTLKKNIVEE-QTYINKWSDSLTNL--TAEVDNLSALDGYRADELTQYKNRNER 833
 QY 136 KQEH---SQHREGGTP-----RNDGVALARSQGRYTTD---DGYIFNASDIIE-- 178
 DB 834 KEKFLSTLKEQEDDIPDGKNIYEYNNHKNVMVKNKEHKISSDINQCNENIIEKKNLETF 893
 QY 179 ----DTGDYIVPHGDHYHYPK-----NELSASELAABAFUSGRGNLSNRYR-RQ 227
 DB 894 NTLVQTLDHAHTGKKQKHVLDLQKPTNLEKLNLELESFGKSLNGSASTTNKQIENIRK 953
 QY 228 NSDNTSRTNWVPSVSNPQTNTNTNNSNTSQAQSDNDISLLKQLYKLPLSORHVESD 287
 DB 954 NIDTIKSLNF-----AKNSSESLSLENIINKKADLIKLDQHTQOELEK- 998
 QY 288 GLVPDPAQITSRTARGVAVPHGDHYHFIPIYSQMSLEERARIIPLRYSNHWVPSDRPE 347
 DB 999 -----HTFIENEEMSP-----LSVIKKEKN----- 1019
 QY 348 QPSQPTPEPSGPQAPNPKIDNSLSVQLVKVGEVYFEKIGSRVYFAKDLPSSET 407
 DB 1020 -----RVESDMSEELIKQLNTKINAILEYNNKSKDRFNGDDDTNLEE 1061
 QY 408 VKNLESKLSKQESVSHLTAKENAVPRDQEFYDKAYN-LLTEAHK---ALFYXNKGNSD 463
 DB 1062 LDFPKQCCQAQBIKKLTT-----NPNVLDNGINVIIEQHEKVIILSENHITKED 1113
 QY 464 FOALDKLLERLNDESTNKEKLVDDLLAFAPITHPERLKGKPNSEQIYEYTED---EVRIAQL 520

DB 1114 KKINEKIQNVNSLNMETKL-----GLLKINEDIKNSRDTTITKSKIOEF 1158
 QY 521 ADKYTTSDGYIFDEHDIISDEGDYVTPHMGSHSHWIGKDSLSDEKVAQAAYTEKKGILP 580
 DB 1159 EKKVQT-----IFGSDVANKKIDA-----IKKEHDVNKDEF-DKEKVKDTSFDEKKKSIE 1208
 QY 581 PSPDADVAVKANTGDSAAAIYNRVKGKRIPLVRLPY---MVEHTVEVKNGL-----IIPH 633
 DB 1209 KAYE---KVGNTLKELEKMDDEKNIKEVEBAQIQYKRIFFIDHDVNLNMDVEKSKIVME 1265
 QY 634 KDHYNIKFAWFDHHTYKAPNGYTYLEDLFAITKYVVEHPDERPHSN---DGMGNASEHVL 690
 DB 1266 KIELYKKEI---DEIKQKNEYKQGD---TSNFYITEQYNSATQSKAKIEQINIAATTKK 1319
 QY 691 GKQHSDEPNKNFKADEPVEETPAEPVQVQVETKEVAQLKE-----AEVLAKVTDSLSL 746
 DB 1320 GTSSTSQDINEL---ESIKEEVHKNLQLVQESNSMEEMRKQILSMKDLLILNNSETIA 1375
 QY 747 K--ANATETLAGLRNNL-----TLQIMDNNSIMAEAKL-----LALLKGSNPFSSVSK 792
 DB 1376 KEISNNTONALGFRENAKTKLNTDELLQORVAAMIEEAKAHKNNDIDIALEDAQIDTEVSK 1435
 QY 793 -EKIN 796
 DB 1436 IEQIN 1440

 RESULT 12
 YOH8 YEAST
 ID YOH8 YEAST STANDARD; PRT; 1176 AA.
 AC Q08236;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 131.4 kDa protein in REX4-ATP19 intergenic region.
 GN Y01078W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 RN NCBI_TaxID=4932;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=9178509;
 RA Tzermita M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast chromosome XV reveals eight known genes and ten new open reading frames including homologues of ABC transporters, inositol phosphatases and human expressed sequence tags";
 RL Yeast 13:583-589(1997).
 CC -1- SIMILARITY: BELONGS TO THE SIN1 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; Z74820; CAA99089.1; --
 DR EMBL; Z74819; CAA99088.1; --
 DR PIR; S66771; S66771.
 DR Germonline; 143500; --
 DR SGD; S0005438; Y01078W.
 DR GO; GO:0005737; C:cytoplasm; IPI.
 DR GO; GO:0001558; P:regulation of cell growth; IPI.
 DR InterPro; IPR008828; SIN1.
 DR Pfam; PF05422; SIN1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1176 AA; 131378 MW; ED2B30D014EB5652 CRC64;

 Query Match 3.6%; Score 150.5; DB 1; Length 1176;

		Best Local similarity	18.6%;	Pred. No. 1.7;	Mismatches	118;	Gaps	41;
					Mismatches	254;	Indels	313;
QY	104	GVYKVDGYYVYLKDAAHADNVRTKEENRQKHQSQHREGGTPRNDGAVA---	LARSQ	160				
DB	329	GKIFLTDNK-----NDGQKSDLSLNANKGI-----HGDSSASGNSVS	RDGLTETE	374				
QY	161	GRYTDDGYIFNASDIIE--DTGDAYIVPHGDHYHIKPNELUSASELAABAFUG---	214					
DB	375	NNNISDWESYINEKDLDDLNFDTVTSNI-----NKTVSIDLGHSTINDGTAV	421					
QY	215	-----RGNLNSITYRRONSDN-TSRTNWVPYSNPCTNTNTSNNTNSQAS-----	262					
DB	422	MNRDSKDRSNSNEFNAQRDRITPGSSYGKSLSGSEYEERYNNDSSTMSGEMLDS	481					
QY	263	--OSNDIDSLLQLYKLPIQSOR-----HVESDGL---VFDPAGIITSRTARGVA	VPHGDHY	312				
DB	482	DMQTNTIPS-----HSIPMSQMKGVIYHGDDSDTLANNVDFKAVLTWNSSR-	526					
QY	313	HFIYQSMSELEBRIARIIPLRVRSNHWPDSRPQPSPQTPPEPSPGPAPANLKIDSN	372					
DB	527	-----HPKERRDTVISGKEPTSLTSSNRKFVSVS	554					
QY	373	SSLV---SQLRVKGEGYVEEKGISRYVFAKDLPSETVKNLSEKLSK-QEYS	423					
DB	555	SNLTSTRSPLLRGHGR-----TSTASSEMKAPKVSDSVLHRARKS	596					
QY	424	TLTAKKENVAP-----RDOEFYDKAYNLLTEAHAKA-----LFXNKG	RNSDFQAL	467				
DB	597	TLTLKQDHQSQSPVSSVHKSKEGNILLIKTXTYLVSKPKASQLSNFNFKKKRTNTNSV	656					
QY	468	DKL-----LE-----RLNDEST--NKEKLVDDLLAFAPITHP	498					
DB	657	DLVEYFSFCGDKVPNYESMGLIYIQASKKYKRNSFTTKVRKSSTIFEVIGFALFLYST	716					
QY	499	ERLCKPNSQIETYETEDEVRIAQLAD-----KYTTSDGYIPDEHDIILSDEGD	VVTVMGH	552				
DB	717	EK-KPDN---FPEDGLTVEDI SNPNFSLKI VDEDEGE PEDN-----	754					
QY	553	SHWTGK-----DSLSDREKVAQAAYTKSG---ILPPSP-----DAD-VKAN	PTG	593				
DB	755	---FGKLDKSTIQISDSSEVLWCKVDDAEKSNQEIETPLPFETGGGLMDASTLDANS	SSH	811				
QY	594	DSAAAIYNRVKGEKRIPLVRLPYMVHTVEVKGNGNLIIPH KDHYHNLIKFAWFDDHTYKAP	653					
DB	812	DTTDGTTINQLSFYK-PIIG---NEDDIDKTNISKII-----DVTVLYP	851					
QY	654	N-----GYTTLEDLPATIKYVVEHPDERPHSNDGWGNASEHVLGK--KDHS	EDFN	700				
DB	852	NVNPKFYNTTISLVAT-----SHIND-----ILLVKYCMKNMDEYALKV	892					
QY	701	--KNF-----KADEBPVEETPAPEPVOVETEKVEAQLEAVEV-----LLA	739					
DB	893	LGKNYILDNDTVLRLDGINKVELISKDOARELHEKOPDLKPVLPTTQSNDLTPTLTL	952					
QY	740	KVTDSSLSKANATETLAGLRNLLTIQIMDNNSIMAAEKL-----LALLKGSN	PSVSSKEK	794				
DB	953	EPLNSYLIKADGAGVAA-----IPENTKVTSKAKISTKYKGLGLAKOHSSSSVASGS	1004					
QY	795	IN	796					
DB	1005	VS	1006					

```

GN PTPN21 OR PTP25.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2E).
RC STRAIN=Sprague-Dawley;
RX MDLINE=951014449; PubMed=7805871;
RA L'Abbe D., Banville D., Tong Y., Stocco S., Ma S.,
RA Fantus G., Shen S.H.;
RT Identification of a novel protein tyrosine phosphatase with sequence
RT homology to the cytoskeletal proteins of the band 4.1 family.";
RL FEBS Lett. 356:351-356(1994).
CC -I- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -I- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62728-1; Sequence=Displayed;
CC Name=2E;
CC IsoId=Q62728-2; Sequence=VSP_000498;
CC -I- TISSUE SPECIFICITY: Particularly abundantly in adrenal glands.
CC -I- SIMILARITY: Contains 1 PERM domain.
CC -I- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC -I- Non-receptor class subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

[illegible]

RESULT 13		
PTNL RAT	PTNL RAT	PTNL RAT
ID	STANDARD:	PRT: 1175 AA.
AC	Q62732;	
Q62728;	Q62732;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Protein tyrosine phosphatase, non-receptor type 21 (EC 3.1.3.48)	
DE	(Protein-tyrosine phosphatase 2E).	

```

Db 261 NMHNKSFFALELANKEETIQOTEDMET-----AKYVMRLCVARHKFYRLNQC 309
QY 131 EINRQKQHSQHQREGTSPRNDGAVALSQ-----GRYTTDDGYIFNASDII 177
Db 310 NLOQAAALNSVRRGSSR---MSLPKPPQVAMPPPQLHYNGHYT-----BPF 355
QY 178 EDTGDYIVPHGDHYHYIPKNELSASAAAAEFLSGRGNLSNRTY--RRONSNTSRT 235
Db 356 ASSQDNVFPVKNKGYCHQSLSLDTQID-----LSGR--IRNGSVYSAHSTNSLNTQP 408
QY 236 NWVPS--VSNPQTTWNTNSNTNSQASQ-----SNDIDSLKQLYK-LPLSORH 283
Db 409 YLOPSMSSNPSPGSDVMRPDIYPSHRHSALIPPSYRPTDYESVMKRLNRCMVHADRH 468
QY 284 VES-----DGLVDFDAQITSRARGVAVPHGDHYH-----IPYSQ 319
Db 469 SHSLNLSIGSYAYSRPDALVYSQPEI--REHPLHASPQSAHYFNLNYSFHSQAPYPY 526
QY 320 MSELERARIIPLYRNSHNVDPDSRPE-----QPSQPT-----PEPSGPQ 362
Db 527 PVERRPVVGAV-----SVPELTNVQLQAQDYPAPIIMRTQVYRPPPPYPYPR 573
QY 363 PA---PNLK-----IDNSLSVLQVRKGVGYFEEKGISRYVFAKDLPSSETVKNLESK 414
Db 574 PANSTPDLRSLHYIYSSSNPDLITRRVHSVQ--TFQEDSLPVAHSLSQVSEPLTAARHAH 631
QY 415 LSKQESV-----SHTLTAKKENVAPR-----DOEFYDKAYNLLTEAH- 451
Db 632 LQKNSNTEIAGLTGFEGLRUKETMSASAADVAPRTFSAGSQSVFSDKVQEGTEBQ 691
QY 452 KALFANXKGRNSDFQAL---DKLLRLNDESTNKEKLVDDL-----AFLAP 494
Db 692 SGGYSHKSLSDATMLIHSSEDELEDDSSREHAVSEPLTAAPFSQQLNYPCASVTP 751
QY 495 ITHPERLGKPNQSYEYDEVRIAQLADKYTTSQGYIPDEHDIISDEGDYVTPHMGHSH 554
Db 752 VTGELHIFEPKSHV--TEPEKRAKDIPVHLVME-----THQPRRH 790
QY 555 WIGKDSLSDEKVAQAAYTKEKGILPPSPDADVANKPTGDSAAALYNVKG 605
Db 791 GLLTFSSESGLTTSGRYARRDSL-----KKRPVSDLLSGKKNVTEG 833

RESULT 14
ID RBPI PLAVB STANDARD; PRT: 2869 AA.
AC Q00738;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPI.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
ON NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
  merozoites."
RL Cell 69:1213-1226(1992).
CC CC -|- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC CC human reticulocyte cells.
CC CC -|- SUBUNIT: Homodimer (Potential).
CC CC -|- SUBCELLULAR LOCATION: Membrane-bound.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

or send an email to license@isb-sib.ch).
CC EMBL; M88097; AAA29743.1;
CC Malaria; Receptor; Signal; Transmembrane.
KW SIGNAL 1 17
FT CHAIN 18 2869
FT DOMAIN 18 2807
FT TRANSMEM 2808 2826
FT DOMAIN 2827 2869
FT SITE 1030 1032
FT SITE 2599 2601
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCEFF CRC64;

Query Match 3.6%; Score 149; DB 1; Length 2869;
Best Local Similarity 17.2%; Pred. NO. 7.1;
Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

QY 24 KOATQKTENLTPDEVSKREGINAEQ-----IVIKITD-----QGY---VTSH 62
Db 1702 KESYEKNLETQVQNSMR---INVEEGSLTDDKKITDIENDLLMKMKQYEEGLQKIKEN 1758
QY 63 GDHYHYTYNGKVPYDAIISELLMKDPN-----YKLDEDIVNEKGVGVKVDGKYVY 116
Db 1759 AD-----KRKSNFELVGSSEINALLDPSTSIKILKKEYDMTGLK-NYGVMMIEHGEF 1812
QY 117 LKD-----AAHADN-----VRTKEINRQKQHSQHRE----- 144
Db 1813 TKSYNLIETHLSNATDYSVTPEKAQSLRELAEEKEEHLRRREBEAIFLLNDIKKVESLKL 1872
QY 145 -----GTPRNDGAVALARSGRVTDD-----GIFNASDIIDETG 181
Db 1873 LKEMMKVSAEYEGMKDHTSVSQLVQDMKTVLDELKTLNDISECSSVLNNVSVIVKVK 1932
QY 182 DAYIVPHGDHYHYIPKNELSASAAAAEFLSGRGNLSNRTYRQNSDNTS---RTNW 238
Db 1933 ES---KHADYRR--DANSWYESMTYLANFLSLDEAKISSGMEFNAEMKSNFKTLELEIF 1987
QY 239 PSVSNPQTTWNTNSN-----TNSQASQNSNDISLLKQLYKLPQSRHVESDGLV 290
Db 1988 SVISNELLKKIPQSDNDVIQKERESEQLAKADATDIYVVK--LKNEFNEKLEAKNKE 2045
QY 291 FDPAQITSRARGVAVPHGDHYHYIPYSQ-----SELEERTARIIPLYRNSHNVDPDSRPE 347
Db 2046 EVSEKVRKALKRLSQVQEGIRCFHFNRLLDNTTELENLKKMVTI-YR-----DKSE 2098
QY 348 QPSPQTPPEPSPGPAPNKLKIDSNSLSVLQVRKGVGYFEEKGISRYVFAKDLPSSET 407
Db 2099 RES-----GLQEMEN-EMNTYSNSITQL-----EGIVVSAGESKEDIKLERNEE 2143
QY 408 VKNLESKLSQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGKNSDFQAL 467
Db 2144 MRNISEKIS-----TIDSKVIEMNSTIDELYKLGKN--COAHWISLISYTA--MKT 2191
QY 468 DKLLERLNDSTNKEKLVDDLAPLAPITHPERLGKPNQSYEYDEVRIAQLADKYTTS 527
Db 2192 SKLLIMINKENKTEKCD-----YKONSST 2219
QY 528 DGYI-----FDEHDIISDEGDYVTPHMGHSHWIGKDSLSDEKVAQAAYTKE 575
Db 2220 DGYVETLKGFGYKSLTFSSASEIVQNAATYSVNFAGHE-----KESLNAIRDIKKELYLPH 2275
QY 576 KGILPSPDADVANKPTG-DSAAAIYNRVKGEKR-----IPLVRLPVWVHEHTVEV-- 624
Db 2276 Q-----NSDISIVEGGVQNMALYDKLNEEKREMDYLNISSETKQKQ-EHSTDFVK 2327
QY 625 -----KNGNLIIP----- 632
Db 2328 PMIELHKGMMETNNKSLLEKELKSLVNDHMSMEAMIKKNGLKYTPESVQNNIYSVI 2387
QY 633 -----HKDHY-----HNKKFAWFDHDT-----YKAPNYTLEDLFATI 665
Db 2388 EAEVKTLEIDRDYGDNYQIVVEHKKQFSLIDRTNALMDDIDEIFKKNYNNL----- 2440

```

QY 666 KYVHEPDERPHS-NDGWNASHVHLGKDHSEDPNKNFKADEPVEETPAEP----- 717

Db 2441 ---MEVNTETHRVNDYIEKTNKLVQAKTEYEQILENIKQNDMDLQIFLKVSIIIFYF 2497

QY 718 EYVQVETKVEAQLAEVLLAKVTD--SSLKANATETLAGLRNLLTIQIMDNNSIMAEA 775

Db 2498 ENVKKKESILNDLYEQERLL-KIGEHLDKEIKNVTEILSSVEIDQKEMMSKN-LLEKK 2555

QY 776 EKLL 779

Db 2556 SKMM 2559

RESULT 15

ID CUT7 SCHPO STANDARD; PRT; 1085 AA.

AC P24339;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Kinesin-like protein cut7.

GN CUT7 OR SPAC25G10.07C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RP MEDLINE=91015362; PubMed=2145514;

RA Hagan I., Yanagida M.

RT "Novel potential mitotic motor protein encoded by the fission yeast cut7+ gene."

RL Nature 347:563-566(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.

RT "The genome sequence of Schizosaccharomyces pombe."

RL Nature 415:871-880(2002).

CC -1- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM G2 TO M PHASE OF THE CELL CYCLE, THE SPINDLE POLE BODY DUPLICATES; THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO FORM A SHORT SPINDLE THAT ELONGATES TOWARD THE NUCLEUS AT METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. B1MC SUBFAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X57513; CAA40738.1; -.

DR EMBL; Z70691; CAA94636.1; -.

DR F1R; T38378; T38378.

DR HSP; F17119; 3KAR.

DR GenDB SPombe; SPAC25G10.07c; -.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.

KW Motor protein; Cell division; Microtubule; ATP-binding; Coiled coil;

KW Mitosis; Cell cycle; Phosphorylation; Repeat.

FT DOMAIN 70 435 KINESIN-MOTOR (BY SIMILARITY).

FT DOMAIN 436 604 COILED COIL (POTENTIAL).

FT DOMAIN 715 740 COILED COIL (POTENTIAL).

FT DOMAIN 897 955 COILED COIL (POTENTIAL).

FT NP_BIND 159 166 ATP (BY SIMILARITY).

FT REPEAT 987 998

FT REPEAT 999 1010

FT MOD_RES 1011 1011

FT PHOSPHORYLATION (BY CDC2) (BY

FT SIMILARITY)

FT SASNPKRKEPTTIDTGYDSDTNSPT ->

FT LRAILGNDVSLLLTL (IN REF. 1).

FT

SQ SEQUENCE 1085 AA; 122133 MW; 566927787559D58 CRC64;

Query Match 3.5%; Score 145.5; DB 1; Length 1085;

Best Local Similarity 18.6%; Pred. No. 2.8; Indels 301; Gaps 41;

Matches 182; Conservative 136; Mismatches 357;

QY 23 GKQATQKTENLTPDEVSKREGINAE-----QIVIKI--TDQYVTSVSHGHYHYN 70

Db 159 GQTGCTKTYTMSGD-LSDSGILSEGAGLIPRALYQLFSSLDNSNQYAVK-CSYVELYN 216

QY 71 GKVPYDAITSEILLKMDPNYKLUDED-----IVNEKGGVIVKVDGYVYLKDAHADN 125

Db 217 EEI-RDLLVSEEL--RKPAPVFDTSRRGNVVITGIEEY-IKNAGDGLRLREGSHRRQ 272

QY 126 VRTKEINQKQEHSS-----QHREGTTPRDGAVALARSGRYTTDDGYIFNASDIIEDT 180

Db 273 VAATKCNLSSSHSIFTTILHRKVSNGMTDNTSLINN-----NSDILLR-- 319

QY 181 GDAYIVPHGDHYVYIPKNELASLAAAFILSGRGNLSNRT----- 223

Db 320 -----ASKLHMVDLAGSENI--GRSGAENKRARETCGMINQSILLTLGRVI 361

QY 224 -----YR-----RONSNTSRTNWVPVSNPPTNTNTNNSNTNSQASOS 264

Db 362 NALVEKAHPIYRESKLTLLQDSLGKTKTSMITVS---STNTLEETISLEYAARA 418

QY 265 NDIDSLLKLYKLPLSRHVESDGLVDPDAQI-----TSRTARGVAVPHGDHYHFIYSQ 319

Db 419 K----SIRNKPQNQLVFRKVLKDLVLDIERKNDLNATKNGVLAESTYKELMDRVQ 475

QY 320 MSEL-----EERIIARIIPLYRSNHWVPSRPEQSPQTPPEPSGPQAPNLKI----- 369

Db 476 NKDLLQCEQARKLEVLDLVNKS-----SREQLQYVSKSNQEHKKEVEALQLQLVNSSTE 529

QY 370 -----DSNSSLVSQLYRKVCEGYVFEKGLSRVVEAKDLFSETVKNLESKLSQESVSH 424

Db 530 LESVKGSEKLNELVLEIEKRRKYEAKITTVATDLQQYRESKEYIASLYEKLDFT 589

QY 425 LTAKENVAPRDQEFYDKAYNLL-----TEAH 451

Db 590 ERNNKEN-----ENFNWLNKFNLLTLRLSPHSGFTDETNGYFTLLINDFNASMEELLNTHSN 645